

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2002, 07:38:37 ; Search time 133 Seconds
(without alignments)
5514.763 Million cell updates/sec

Title: US-09-845-721-1
Perfect score: 1881
Sequence: 1 gttatttctcaaaaggaaa.....aaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published_Applications_NA:*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	72	3.8	416	10	US-09-960-352-4584
C 2	66.2	3.5	419	10	US-09-960-352-11234
C 3	59.6	3.2	406	10	US-09-960-352-10265
C 4	58.4	3.1	393	10	US-09-960-352-4582
C 5	58.4	3.1	2000	9	US-09-938-842A-2868
C 6	57.4	3.1	337	10	US-09-960-352-6976
C 7	57	3.0	377	10	US-09-960-352-7419
C 8	56.2	3.0	375	10	US-09-960-352-15014
C 9	54.4	2.9	277	10	US-09-960-352-12673
C 10	54.2	2.9	424	10	US-09-960-352-11218
C 11	53.4	2.8	241	10	US-09-960-352-5777
C 12	52.8	2.8	298	10	US-09-960-352-1004
C 13	52.8	2.8	302250	10	US-09-962-832-154
C 14	52.4	2.8	277	10	US-09-960-352-12673
C 15	52.4	2.8	419	10	US-09-960-352-11234
C 16	52	2.8	5332	10	US-09-801-861-3
C 17	51.8	2.8	375	10	US-09-960-352-15014
C 18	51.2	2.7	416	10	US-09-960-352-4584
C 19	50.4	2.7	4877	10	US-09-834-975-872

C 20	50.4	2.7	4877	10	US-09-834-975-873	Sequence 873, App
C 21	50.4	2.7	4877	10	US-09-834-975-884	Sequence 884, App
C 22	50.2	2.7	380	10	US-09-960-352-9335	Sequence 9335, App
C 23	50.2	2.7	1797	10	US-09-729-674-89	Sequence 89, Appl
C 24	50	2.7	408	10	US-09-960-352-6263	Sequence 6263, App
C 25	49.8	2.6	397	10	US-09-960-352-13784	Sequence 13784, A
C 26	49.6	2.6	1371	10	US-09-764-853-227	Sequence 227, App
C 27	49.4	2.6	358	10	US-09-960-352-5559	Sequence 5559, App
C 28	49.4	2.6	390	10	US-09-960-352-1976	Sequence 1976, App
C 29	49.4	2.6	640681	10	US-09-790-988-1	Sequence 1, Appll
C 30	49.2	2.6	932	12	US-10-078-929-93	Sequence 93, Appl
C 31	49.2	2.6	1377	10	US-09-925-299-93	Sequence 2, Appll
C 32	49.2	2.6	1931	10	US-09-748-537-2	Sequence 1, Appll
C 33	49.2	2.6	1931	10	US-09-728-721-1	Sequence 2, Appll
C 34	49.2	2.6	1931	12	US-10-133-780-2	Sequence 1, Appll
C 35	49.2	2.6	1931	12	US-10-105-931-1	Sequence 1, Appll
C 36	49	2.6	241	10	US-09-960-352-7904	Sequence 7904, App
C 37	49	2.6	368	10	US-09-834-975-451	Sequence 451, App
C 38	49	2.6	425	10	US-09-834-975-451	Sequence 47, Appl
C 39	49	2.6	2645	10	US-09-764-870-76	Sequence 76, Appl
C 40	48.8	2.6	2000	9	US-09-938-842A-3939	Sequence 3939, App
C 41	48.6	2.6	376	10	US-09-960-352-5087	Sequence 5087, App
C 42	48.6	2.6	302250	10	US-09-962-832-154	Sequence 154, App
C 43	48.4	2.6	415	10	US-09-969-373-751	Sequence 751, App
C 44	48	2.6	2890	10	US-09-925-300-763	Sequence 763, App
C 45	48	2.6	180557	12	US-10-003-806-6	Sequence 6, Appll

ALIGNMENTS

RESULT 1

US-09-960-352-4584/c

; Sequence 4584, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Byatt, John C.

; APPLICANT: Mathalagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 4584

; LENGTH: 416

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 20-LIB3057-016-Q1-K1-E11

US-09-960-352-4584

Query Match

Best Local Similarity 48.3%; Score 72; DB 10; Length 416;

Mismatches 201; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 1466 AACACAAAAAATACCTTTGAAGATAGTGAATAAGAGAAAAAGCTTTAGTGCCTCAGGT 1525

Db 416 AAAAAAATTTTTTTTTTAAATAAATTTTTTATATAAATTTTTTAAATTAATAAT 357

QY 1526 TGTACACAGACTAGAGAAAGCTCAGTTTCACCAATCCACATCAATGAGTTTAAT 1585

Db 356 TACCATATATATTTTAAATATATTTTAAATAATCTTTTATATATCTCTATAAT 297

QY 1586 TTAATTTGTAATAACTGATATTTACTGCCAAATATAAGAAAAATATTTAAGTATTGGTTA 1645

Db 296 TAATAAATTTAAATTTTAAANAATTTTCTTTAAANAATTTTAAANAATTTTAAAA 237

QY 1646 TGGTTGTAATTTCAATGTGAATGCTAATTAGATAGTGCATATATATCAATTTCTTCA 1705

Db 236 TATTTTTTTTATGAATTTTTTAAATTTTATATATATTTTAAATTTTAAATCTTTTT 177

Qy	1706	TTACTTTAACTGATTTGTTGCATGGCAGTTTGGTTAAAGTACTATCAATCAATGTTATATTTGTC	1765
Db	176	TTTTTTTTTTTTTTTTTTTAAATTTTATTTTTTTTTTTTATATAAAAAATTTATTTTTTATT	117
Qy	1766	AAATATTATGTCACACAGAAATATTCNTCTAAGTCTATATTTTTTAAAGGAATAAATACATA	1825
Db	116	ATTTTCTATTATTTTCATTTAAATTTTTTTTTTTTTTAAAAAAAATAAAAAAAATACCAA	57
Qy	1826	GCCTTAAACAGTGTAACTTTTAAATGTAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1881
Db	56	AACCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1

RESULT 2

US-09-960-352-11234/c

; Sequence 11234, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LAC

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960.352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 11234

; LENGTH: 419

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8

US-09-960-352-11234

```

Query Match          3.5%; Score 66.2; DB 10; Length 419;
Best Local Similarity 51.2%; Pred. No. 0.0025;
Matches 155; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 1579 TTTAAATTAAATGCTAAAAAGTGTATTTACTGCCAAATATATAAGAAAAATATTTTAAAGTA 1638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 406 TATATATTTAAAAAATAAAATTTATTTATATAAAAAATATTTTATAAAAATATTTTTT 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1639 TTGGTTAFGTTGTAATTTTCAATGTGAAATGCTAAATTAGATAGTCATATATATCAAT 1698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 346 AAAAAAATTTTTTTTTTAAAAATTTTTTATTATTATTATTATTTTAAATTTATAATTTTTTAAAAA 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1699 TTCCTTCATTACTTAATGCTATTTGTTGCCATGGCAGTTCTTAAAGTACTATCAATGCTATA 1758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 286 ATTTTAAAAATTTTAATTTTAAAAATTTTAAAAAATAATAATAAAAAATTTTTTTTTTATATT 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1759 TTTTGTCAATATTATGTCACACAGAAAAATATTCATGTAGTCATATTTTTTAAGGAATAA 1818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 TAATAAAATTTATAAATTTTAACTTTTTTAAAAATTTTTTTTATTATTATTAAAAAATAA 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1819 ATACATAGCCCTTAAACACAGTGTATAACCTTTAAATGTAAAAAATAAAAAAATAAAAAA 1878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 166 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1879 AAA 1881
    |||
DB 106 AAA 104

RESULT 3
US-09-960-352-10265/c
: Sequence 10265, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengbing
: APPLICANT: Wyatt, John C.
: APPLICANT: Mathaladan, Nagabban

```

```

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10265
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 44-LIB3058-038-Q1-K1-C12
US-09-960-352-10265

```

Query Match	3.2%;	Score	59.6;	DB	10;	Length	406;
Best Local Similarity	49.4%;	Pred. No.	0.034;				
Matches	155;	Conservative	0;	Mismatches	159;	Indels	0;
Qy	1568	TTCAAATGAGT	TTTAAATTTAAATTTGTTAAACCTGATATTA	CTACTGCCAAATATAGA	AAAAA	1627	
Db	405	TTTAAATTAAT	CAAAATATTTTAAATGTCATAACTTCTTTAAATATGATATATTTATTTATAT		346		
Qy	1628	TATTTTAAAGT	ATTGGTTATGTTGTAATTTTCAATGTGAAATGCTAAT	TAGATAGGTCAT	1687		
Db	345	TTAGATATCACT	ATAATAATCTCATGTTAAATCATATATATATGTAAGTTTGTATCAT		286		
Qy	1698	ATATATTCAA	ATTTCCTTCATTAAGTATTTGGTTCGATGGCAGTTTGGTTAAAGTACTA		1747		
Db	285	ATACATTTTC	ATTTTTTTTATTTTAAATATATGTTGTTCAITTTTATAGTATTTTATATTTAAAT		226		
Qy	1748	TCATGTGTAT	ATTTTTGTCAAATATTTATGTCCACAGAAAAATTTCCATGTAAGTCATATTTT		1807		
Db	225	TATTCCTTTAA	AAAAATAGTATTTATTTAAATATATTTTGAAGAAAAATATATAAAAAAAATATAA		166		
Qy	1808	TTAAGGAATA	TAATACATAGCCTTTAAACAGTGTATACTTTTAAAAATGTAAAAA	AAAAA	1867		
Db	165	TAATAATAA	AAAAAATAAAAAATATAAAAAATATAAAAAATATAAAAAATATAAAAAA		106		
Qy	1868	AAAAA	AAAAA	AAAAA	AAAAA	1881	
Db	105	ATAAATAA	ATAA	92			

```

RESULT 4
US-09-960-352-4582/c
; Sequence 4582, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4582
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 20-LIB3057-013-Q1-K1-E7
US-09-960-352-4582

```

[illegible]


```

RESULT 12
US-09-960-352-1004/c
; Sequence 1004, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

```

Query Match	2.8%	Score 52.8	DB 10	Length 302250
Best Local Similarity	52.1%	Pred. No. 2.2		
Matches 165	Conservative 0	Mismatches 149	Indels 3	Gaps 2
Qy 1567	ATTCAAATGAGCTTTTAAATTTAAATTTGTAAGAACTGATATTACTGCCAAATATAGAAAA	1626		
Db 300411	ATATATAAATGCTTTATATATATTTTATATAAATATTTTAAATATATAATATATATATATTT	300352		
Qy 1627	ATATTTTAAAGTATGGTTATGTTGTAAATTTTTCATGTGGAATGCTAATTAGATAGGTC	1686		
Db 300351	ATATAAATATATATTTTA	300292		
Qy 1687	TATATATCAATTTCTTCAT-TACTTAATGATTTTGGTCATGGCAGTTTGGTTAAAGTAC	1745		
Db 300291	TATATAATATATATTTATATATATATTTTATATATATATATATAACATATATTTTATATA	300232		
Qy 1746	TATCATGTGTATATTTTCTGCA--ATATATTGTCACAGAAATATTTCATGTAACTCAT	1803		
Db 300231	TATTTATTTATATATATATATATATTTTATATATATTTATATATATATAGTATGTATTTATATA	300172		
Qy 1804	TTTTTTAAGGAATAAATACATAGCCTTAAAAACAGTGTAACTTTTAAAAATGTAAAAAAA	1863		

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2002, 07:35:27 : Search time 2401 seconds
(without alignments)
12687.926 Million cell updates/sec

Title: US-09-845-721-1
Perfect score: 1881
Sequence: 1 gttatttcttcaaaaggaaa.....aaaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estha:*
2: em_esthum:*
3: em_estinu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hctc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hctc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES											
Result	No.	Query		Match	Length	DB	ID	Description			
		Score	%								
c	1	609.4	32.4	684	17	B56597		B56597 CIT-HSP-200			
	2	576.2	30.6	761	12	B824968		B824968 C0507E09-			
	3	552.6	29.4	938	12	BF968206		BF968206 602268947			
c	4	541.2	28.8	984	12	BG295999		BG295999 602395255			
	5	497.8	26.5	668	12	B824971		B824971 C0507E12-			
	6	474	25.2	631	12	B824600		B824600 C0501H07-			

7	471	25.0	647	10	BB171739	BB171739
c	462.6	24.6	570	17	AZ494191	AZ494191 IM0329L18
9	451.8	24.0	519	12	BG086308	BG086308 H3124D01-
c	435.8	23.1	587	12	BE824614	BE824614 C0502803-
c	435	23.1	591	12	BE824967	BE824967 C0507E08-
c	431.6	22.9	587	12	BE824766	BE824766 C0504E03-
c	398.4	21.2	559	12	BF075992	BF075992 225541 MA
c	379	20.1	422	9	AL600185	AL600185 DFE2P313J
c	362.6	19.3	520	10	AW061316	AW061316 UT-M-BH1-
c	346.4	18.4	488	12	BE824969	BE824969 C0507E10-
c	340.6	18.1	465	17	AQ222591	AQ222591 HS-2007_B
17	326	17.3	345	14	Z18870	Z18870 HSDHEG001 S
19	320.6	17.0	463	9	AA500806	AA500806 v901e07.r
c	319	17.0	459	12	BE824472	BE824472 C0500A07-
c	313.6	16.7	460	12	BE824814	BE824814 C0505C01-
c	296.8	15.8	461	14	R59799	R59799 yho7h07.r1
c	289.6	15.4	509	14	R61341	R61341 yho7h07.s1
c	244.4	13.0	258	14	R58357	R58357 G3141 Retal
c	240.8	12.8	682	10	BB366337	BB366337
c	223.6	11.9	568	17	AQ608397	AQ608397 HS-2122-B
c	213.6	11.4	333	9	AV147896	AV147896 AV147896
27	184	9.8	266	9	AI614903	AI614903 v901e07.y
c	180.4	9.6	304	10	AW122633	AW122633 UT-M-BH2
c	177	9.4	284	10	BB310100	BB310100 BB310100
c	172.2	9.2	348	10	BB259564	BB259564 BB259564
c	167.8	8.9	330	10	BB131127	BB131127 BB131127
c	167.4	8.9	320	10	BB464297	BB464297 BB464297
c	162.6	8.6	281	14	T20318	T20318 EST009 ED18
c	161.6	8.6	261	10	AV329383	AV329383 AV329383
c	161	8.6	337	10	BB314194	BB314194 BB314194
c	152.6	8.1	235	10	BB176012	BB176012 BB176012
c	149.6	8.0	1012	17	CNS02MDL	AL203970 Tetraodon
c	147	7.8	261	10	BB191298	BB191298 BB191298
c	146.6	7.8	717	9	AI641710	AI641710 fc22h07.x
c	144.6	7.7	235	10	AV341872	AV341872 AV341872
c	142.6	7.6	387	17	B38040	B38040 HS-1048-B2-
c	139.2	7.4	239	10	BB269323	BB269323 BB269323
c	137.6	7.3	290	10	BB188782	BB188782 BB188782
c	134.2	7.1	250	10	BB431762	BB431762 BB431762

ALIGNMENTS

RESULT 1
B56597
LOCUS CIT-HSP-2007F6.TFC CIT-HSP Homo sapiens genomic clone 2007F6, DNA
DEFINITION B56597 684 bp linear
ACCESSION B56597
VERSION B56597.1 GI:2610931
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 684)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
and Venter,J.C. K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
BUILDING Unpublished (1997)
JOURNAL Other GSSs: CIT-HSP-2007F6.TF
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .684
/organism="Homo sapiens"
/db_xref="CDB:7040000"
/db_xref="taxon:9606"
/clone="2007F6"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 222 a 129 c 94 g 239 t

ORIGIN

Query Match 32.4%; Score 609.4; DB 17; Length 684;
Best Local Similarity 99.8%; Pred. No. 3e-74; Indels 0; Gaps 0;
Matches 610; Conservative 0; Mismatches 1;

Qy 1 GTTATTTCTTCAAGAGAAACACAAATTTCTTTATATCAAAACAATGCAAACTTGATG 60
Db 74 GTTATTTCTTCAAGAGAAACACAAATTTCTTTATATCAAAACAATGCAAACTTGATG 133
Qy 61 GTTCTTAATTTCTACATTTCTTATTAATAGTTTACAACTTAAATTAAGTACAC 120
Db 134 GTTCTTAATTTCTACATTTCTTATTAATAGTTTACAACTTAAATTAAGTACAC 193
Qy 121 AATTGAAGATTTTCTTACAAACACACGTTATACGTCATTTAAATGCCAATA 180
Db 194 AATTGAAGATTTTCTTACAAACACACGTTATACGTCATTTAAATGCCAATA 253
Qy 181 TCAATAGTTTATTTCTTATTTCTTCTAGGGAAAAACCAACTGCTCCAAAAGAATGT 240
Db 254 TCAATAGTTTATTTCTTATTTCTTCTAGGGAAAAACCAACTGCTCCAAAAGAATGT 313
Qy 241 GTTTTCTCCGATTTGGAATCAACATGCAAGTCGTAATCTAACATTCACAGTCGAGATG 300
Db 314 GTTTTCTCCGATTTGGAATCAACATGCAAGTCGTAATCTAACATTCACAGTCGAGATG 373
Qy 301 ACATTGATGACATCAACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAG 360
Db 374 ACATTGATGACATCAACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAG 433
Qy 361 TGTCTCTCACCAGGATTTCTTATGTAGAAATTTGTTGGGACTTGGCAGCAACCTCACTG 420
Db 434 TGTCTCTCACCAGGATTTCTTATGTAGAAATTTGTTGGGACTTGGCAGCAACCTCACTG 493
Qy 421 TATTGGTACTTTACTGCAATGAATCCAACTTAATCAACTCTGTCAGTAAACATTAATACAA 480
Db 494 TATTGGTACTTTACTGCAATGAATCCAACTTAATCAACTCTGTCAGTAAACATTAATACAA 553
Qy 481 TGAATCTTCATGTAATGATGTAATTAATTTGTTGGGATGATATTCCTTAAGTATAGTTA 540
Db 554 TGAATCTTCATGTAATGATGTAATTAATTTGTTGGGATGATATTCCTTAAGTATAGTTA 613
Qy 541 TCCCTCTCTTCTACTGGAGAGTAACTGCTCTCATTTGCTTGTGTTCCATGAGGCTTGTG 600
Db 614 TCCCTCTCTTCTACTGGAGAGTAACTGCTCTCATTTGCTTGTGTTCCATGAGGCTTGTG 673
Qy 601 TATCTTTTGTCA 611
Db 674 TATCTTTTGTCA 684

RESULT 2
BE824968/c 761 bp mRNA linear EST 21-SEP-2000
LOCUS BE824968
DEFINITION C0507E09-3 NIA Mouse E13.5 VMB Dopamine cell cDNA Library Mus
musculus cDNA clone C0507E09 3', mRNA sequence.
ACCESSION BE824968
VERSION BE824968.1 GI:10257202

KEYWORDS
SOURCE

house mouse.
Mus musculus

ORGANISM

REFERENCE
AUTHORS

1 (bases 1 to 761)
Barrett,T., Xie,T., Piao,Y., Dillon-Carter,O., Kargul,G.J., Lim
M.K., Chrest,F.J., Wersto,R., Rowley,D.L., Juhaszova,M., Zhou,L.,
Vawter,M.P., Becker,K.G., Cheadle,C., Wood,W.H. III, McCann,U.D.,
Freed,W.J., Ko,M.S.H., Ricaurte,G. and Donovan,D.M.
Genes expressed by midbrain dopamine neurons: a dopamine
neuron-specific cDNA library and microarray used to examine changes
in gene expression during development and methamphetamine
neurotoxicity
Unpublished (2000)
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6920, USA
Email: cdna@lgsun.grc.nia.nih.gov
niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)
Plate: C0507 row: E column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 761
POLYA=Yes.

TITLE

JOURNAL
COMMENT

Unpublished (2000)
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6920, USA
Email: cdna@lgsun.grc.nia.nih.gov
niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)
Plate: C0507 row: E column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 761
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1. .761
/organism="Mus musculus"
/strain="TH-beta-gal transgenic mouse"
/db_xref="niaEST:C0507E09-3"
/db_xref="taxon:10090"
/clone="C0507E09"
/clone_lib="NIA Mouse E13.5 VMB Dopamine cell cDNA
Library"
/sex="Male/Female"
/dev_stage="13.5dpc"
/lab_host="DH10B"

15'-pGAGTATTCTAGATCGGCGGCCCTTTTCTTTTCTTTT-3' from
0.91ug of total RNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker L1-Sal4 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from free
linkers by Centricon 100. Then, cDNAs were amplified by
long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Yulan Piao and Minoru.S. H. Ko."

BASE COUNT 244 a 139 c 128 g 250 t

ORIGIN

Query Match 30.6%; Score 576.2; DB 12; Length 761;
Best Local Similarity 87.4%; Pred. No. 9.3e-70;
Matches 567; Conservative 0; Mismatches 88; Indels 8; Gaps 3;

Qy 1074 GTAGTCTTTTGTGTAAGAACTTCAGTTCTGTGTAATAATTGCCCTCCGCGAGCTGTGAAA 1133
Db 761 GTCTGTTTGTGTTGTAAGAACTTCAGTTCTGTGTAATAATTGCCCTCCGCGAGCTGTGAAA 702
Qy 1134 CGACACCGTGNACGACGAGAAAGACAAAGAGAGTCTTCAGGATGTCCTTATTGATTATT 1193
Db 701 GCCCACCGGACCGACGAGAAAGAGTCTTCAGGATGTCCTTATTGATTATT 642


```
Qy 1194 TCTACATTTCTCTCTGCTGACACCAATTTCTGTTTAAATACCACCAATTTATGTTA 1253
Db 1194 TCTACATTTCTCTCTGCTGACACCAATTTCTGTTTAAATACCACCAATTTATGTTA 582
Qy 1254 GGCCCAAGTGAACCTTTAGTAAATTAAGATTTGTTTATGTCATGGCTTTATGGAACA 1313
Db 1254 GGCCCAAGTGAACCTTTAGTAAATTAAGATTTGTTTATGTCATGGCTTTATGGAACA 522
Qy 1314 ACTATATTTACCCCTCTATATATGATTCACATGACAAATTTCAAAGGCTTGAAGA 1373
Db 1314 ACTATATTTACCCCTCTCTGTTATGATTCACACGAAAGTTCCAAAAGGCTTGAAG 462
Qy 1374 AGTAAATGAAGGAGGTTGTTTCTATAGTACAGCTGATCCCTGCCCTAAATAGCT 1433
Db 1374 AGTAAATGAAGGAGGTTGTTTCTATAGTACAGCTGATCCCTGCCCTAAATAGCT 402
Qy 1434 GTAATACACAACTCTTGGATAGATCCCAAGAAACAAAATTTACCTTTGAAGATAGT 1493
Db 1434 GTAATACACAACTCTTGGATAGATCCCTAAAGAAACAAAAGGTTACCTATGAAGACAGT 342
Qy 1494 GAAATAGAGAAAACGTTTGTAGTCCCTCAGGTTGTCAGACTAGAGAAAAGTCTCAGTT 1553
Db 1494 GAAATAGAGAGAAATGTTTGTAGTACCTCAGGTTGTCAGACTAGAGAAAAGTCTCAGG 282
Qy 1554 TCACCAATCCACATTCACATGAGTTTAAATTTAAATTTGTAAGAACTGATATTACTGCC 1613
Db 1554 TCACCAATCCACATTCACATGAGTTTAAATTTAAATTTGTAAGAACTGATATTACTGCC 224
Qy 1614 AAATATAGAAAATATTTTAAATTTGTTTAAATTTGTAAGAACTGATATTACTGCC 1673
Db 1614 AAATATAGAAAATATTTTAAATTTGTTTAAATTTGTAAGAACTGATATTACTGCC 164
Qy 1674 ATTAGATAGGTCATATATATATATATATATATATATATATATATATATATATATAT 1729
Db 1674 ATTAGATAGGTCATATATATATATATATATATATATATATATATATATATATATAT 104
Qy 1730 CAGTTTGTGTTTAAAGTATATATATATATATATATATATATATATATATATATATAT 1787
Db 1730 CAGTTTGTGTTTAAAGTATATATATATATATATATATATATATATATATATATATAT 44
Qy 1788 ATTATGTAAGTATATATATATATATATATATATATATATATATATATATATATATAT 1830
Db 1788 ATTATGTAAGTATATATATATATATATATATATATATATATATATATATATATATAT 1

RESULT 3
LOCUS BF968206 938 bp mRNA linear EST 22-JAN-2001
DEFINITION 602268947F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357426 5',
mRNA sequence.
ACCESSION BF968206
VERSION BF968206.1 GI:12335421
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 938)
NIH-MGC http://mgi.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM993 row: m column: 11
High quality sequence stop: 711.
```

```
FEATURES
source
Location/Qualifiers
1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4357426"
/tissue_lib="NIH_MGC_84"
/lab_host="NIH_MGC_84"
/lab_host="adrenal cortex carcinoma, cell line"
/label="Organ: adrenal gland; Vector: PCMV-SPORT6; Site_1:
Noti; Site_2: SalI; Cloned unidirectionally; oligo-dr
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 286 a 208 c 156 g 288 t
ORIGIN
Query Match 29.4%; Score 552.6; DB 12; Length 938;
Best Local Similarity 88.3%; Pred. No. 1.3e-66;
Matches 658; Conservative 0; Mismatches 79; Indels 8; Gaps 5;
Qy 112 TAAGTACAAATTCGAAGATTTTCTTCTTACAAAGAACAGCTTATAGCTTATTAAT 171
Db 112 TAAGTACAAATTCGAAGATTTTCTTCTTACAAAGAACAGCTTATAGCTTATTAAT 171
Qy 171 TAAATCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 171 TAAATCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Qy 172 TGCCAAATATCAATAGTTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 231
Db 172 TGCCAAATATCAATAGTTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 231
Qy 61 CATGAATATCAATAGTTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 116
Db 61 CATGAATATCAATAGTTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 116
Qy 232 AAAGAATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 291
Db 232 AAAGAATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 291
Qy 117 AAAGAATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 176
Db 117 AAAGAATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 176
Qy 292 TGCAGATGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 351
Db 292 TGCAGATGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 351
Qy 177 TGCAGATGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 236
Db 177 TGCAGATGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 236
Qy 352 GCTTTCAAGTCTCTCTCACCAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 411
Db 352 GCTTTCAAGTCTCTCTCACCAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 411
Qy 237 GCTTTCAAGTCTCTCTCACCAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 296
Db 237 GCTTTCAAGTCTCTCTCACCAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 296
Qy 412 ACCCTCAGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 471
Db 412 ACCCTCAGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 471
Qy 297 ACCCTCAGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 356
Db 297 ACCCTCAGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 356
Qy 472 TTATTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 531
Db 472 TTATTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 531
Qy 357 TTATTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 416
Db 357 TTATTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 416
Qy 532 CTATAGTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 591
Db 532 CTATAGTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 591
Qy 417 CTATAGTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 476
Db 417 CTATAGTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 476
Qy 592 AGGCTTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 651
Db 592 AGGCTTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 651
Qy 477 AGGCTTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 536
Db 477 AGGCTTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 536
Qy 652 ACAGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 710
Db 652 ACAGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 710
Qy 537 ACAGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 596
Db 537 ACAGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 596
Qy 711 ATGTTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 769
Db 711 ATGTTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 769
Qy 597 ATGTTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 656
Db 597 ATGTTAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 656
Qy 770 GGTAAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 829
Db 770 GGTAAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 829
Qy 657 GGTCAA -TTTTTTCAGTCTTC -AAGTGAATATCTGGAAGAACACAGACACTTTATGTCGT 714
Db 657 GGTCAA -TTTTTTCAGTCTTC -AAGTGAATATCTGGAAGAACACAGACACTTTATGTCGT 714
Qy 830 CAGTACAAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 854
Db 830 CAGTACAAATGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 854
Qy 715 ACCATTGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 739
Db 715 ACCATTGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 739
```

```
RESULT 4
BG295999      984 bp      mRNA      linear      EST 21-FEB-2001
LOCUS         602395255F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4507188 5',
DEFINITION    mRNA sequence.
ACCESSION     BG295999
VERSION       BG295999.1 GI:13058195
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               1 (bases 1 to 984)
               NIH-MGC http://mgc.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
               Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs@mail.nih.gov
               Tissue Procurement: The Cepko Laboratory
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM10383 row: m column: 13
               High quality sequence stop: 689.
               Location/Qualifiers
                 1..984
                   /organism="Mus musculus"
                   /db_xref="taxon:10090"
                   /clone_lib="IMAGE:4507188"
                   /clone_lib="NIH_MGC_94"
                   /tissue_type="retina"
                   /lab_host="DH10B (phage-resistant)"
                   /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                   Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                   Average insert size 3.3 kb. Library enriched for
                   full-length clones and constructed by Life Technologies.
                   Note: this is a NIH_MGC Library."
BASE COUNT    254 a 236 c 198 g 295 t 1 others
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

```
Db 360 ATGCTAATGACATCCATTTGGATTTTCTTCTTCTCATCTCTGATTCCTCTCATTTGAA 419
Qy 771 GTAAA-TTTTTCAGTCTTCAAGTGAATATCCTGGGAAAACAGACACTTTTATGTCT 829
Db 420 GTAAAGTTTTCAGTCTTCAAGTGAATATCAGTGGGAAAACAGACACTGCTGTGTCT 479
Qy 830 CAGTACAAATGAATACTACACTGAATGCGAATGATTTATACCTCTTCTTATAGTACAGATCC 889
Db 480 CAGTACAAGTGAATACTACTACTGAGCTCGGATGTACTATACCTTTTGGTCAGATCCC 539
Qy 890 AATATCTTTTTCACCTGTTGTAGTAAATGTAATACATACACAAAATACTTCAGGCTCT 949
Db 540 CATCTTCTTCTTTCACAGTTTATAGTCATGTGTATCATACATACATAGTACTCAGGCTCT 599
Qy 950 TAATATCGAATAGGCACAAAGATTT-TCAACAGGGGAGAGAAGAA-----GCAAGAAG 1004
Db 600 TAACATCCCATAGGCATAGATTTCTCAACAGGAGAGAAGAGAAANGCCCGGAAACG 659
Qy 1005 AAAAGACAAATTTCTTAACC-ACACAACATGAGGCTACAGACATGTCTACAA-----AGC 1058
Db 660 AAAAGACAAATCTCTAGCTTACACAATGAAGACCACAGGACATGTCAAAAAGCACT 719
Qy 1059 AGTGGTGGGAGAAATGTAGTCTTTGGTGTGAAGAACTTCAAGTTTCTGTAATATATGCCCC 1118
Db 720 GGGTGGGAGAAATGTGCGGGTTGGGGTGACAACTTCACTTTCTCTGTATAATTTGCCCTC 779
Qy 1119 CGGCG-AGCTGTGAACGACACCGGTGAACGACGAGAGAAAGAGAGAGTCTTCAGGAT 1177
Db 780 CGGGAAGCGGTGAACGACCGGTGACGAGAGACGGGGGAACGAGTTTTCAAAT 839
Qy 1178 GTCTTTATGATTTATTTTACATTTCTTCTCTGCTGGACACCAATTTCTGTTTAA 1233
Db 840 GTTGTATGGA-TATTTCTAAATTTCTTTCCTGGGACACCATCTTGTTAATTA 894
RESULT 5
BE824971/c 668 bp mRNA linear EST 21-SEP-2000
LOCUS C0507E12-3 NTA Mouse E13.5 VMB Dopamine cell cDNA Library Mus
DEFINITION musculus cDNA clone C0507E12 3', mRNA sequence.
ACCESSION BE824971
VERSION BE824971.1 GI:10257205
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 668)
AUTHORS Barrett,T., Xie,T., Piao,Y., Dillon-Carter,O., Kargul,G.J., Lim
,M.K., Chrest,F.J., Wersto,R., Rowley,D.L., Juhaszova,M., Zhou,L.,
Vavter,M.P., Becker,K.G., Cheadle,C., Wood,W.H. III, McCann,U.D.,
Freed,W.J., Ko,M.S.H., Ricaurte,G. and Donovan,D.M.
TITLE Genes expressed by midbrain dopamine neurons: a dopamine
neuron-specific cDNA library and microarray used to examine changes
in gene expression during development and methamphetamine
neurotoxicity
JOURNAL Unpublished (2000)
COMMENT Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
niaEST (http://lgsun.grc.nia.nih.gov/cdna/cdna.html)
Plate: C0507 row: E column: 12
Seq primer: -2LM13 Forward
High quality sequence stop: 668
POLYA=Yes.
Location/Qualifiers
  1..668
    /organism="Mus musculus"
    /strain="TH-beta-gal transgenic mouse"
    /db_xref="niaEST:C0507E12-3"
    /db_xref="taxon:10090"
```

```

/clone="C0507E12"
/clone_lib="NIA Mouse E13.5 VMB Dopamine cell cDNA
Library"
/sex="Male/Female"
/dev_stage="13.5dpc"
/lab_host="DH10B"

Note: Vector: pSPORT1 (Gibco/BRL Life Technology);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
3000 Dopamine cells (cell collected by Dr. Tanya Barrett).
The double-stranded cDNA was synthesized by Gibco's kit
with an Oligo(dT) primer [NotI primer-adaptor from
GibcoBRL]
[5'-pGACTAGTCTAGATCGGAGCGGCCCTTTTCTTTTCTTTT-3'] from
0.9ug of total RNA. The double-stranded cDNAs were
treated with T4 DNA polymerase and purified by
ethanol-precipitation. The cDNAs were ligated to
Eone-linker L1-Sal4 (include SalI sequence). The cDNAs
were purified by phenol/chloroform and separated from free
linkers by Centricon 100. Then, cDNAs were amplified by
long-range high fidelity PCR using Takara's Ex Taq
polymerase. Then, the cDNAs were purified by
phenol/chloroform and by Centricon 100. The cDNAs were
digested with SalI and NotI enzymes. Then, the cDNAs were
size selected by Gibco's Size Fractionation Column. The
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by chemical method. The library was
constructed by Yulan Piao and Minoru S. H. Ko."

BASE COUNT      228 a 109 c 102 g      229 t
ORIGIN

Query Match      26.5%; Score 497.8; DB 12; Length 668;
Best Local Similarity 87.9%; Pred. No. 4.9e-59;
Matches 590; Conservative 0; Mismatches 72; Indels 9; Gaps 34;

QY 1166 AGCTTCAGGATGCTTTATTGATTATTTCTACATTTCTCTCTGCTGGACACCAATTC 1225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 668 AGCTTCAAAATGCGTTATTGATTATTTCTACATTTCTCTCTGCTGGACACCAATTC 609

QY 1226 TGTTTTAAATACCAACCATTTATGCTTTAGGCCCAAGTGACCTTTTAGTAAATTAAGATT 1285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 608 TGTTTTAAATACCAACCATTTATGCTTTAGGCCCAAGTGACCTTTTAGTAAATTAAGATT 549

QY 1286 GTGTTTTTATGTCATGCTTTATGGAACAACATATATTTCACCCTCTATTATATGATTCAC 1345
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 548 GTGTTTTCTAGTCATGCTTATGGAACAACATATTTCACCCTCTCTTTGATGATTCAC 489

QY 1346 TAGACAAATATTTCAAAGGCTCTTGAAAGTAAATGAAAGGCGAGTTGTTCTCTATAGT 1405
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 488 CAGACAAAGTTTCAAAGGCTTTTAAAGGTAAAGATG-AAAAGGCGAGTTGTTCCATAGT 430

QY 1406 AGAAGCTGATCCCTGCTTAATATGCTGTAATACAAACTCTTTGGATAGATCCCAAAAG 1465
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 429 TGAAGCTGATCCCATGCTTAATACGCTGTAATACAACTCATGGATAGATCCTAAAG 370

QY 1466 AAACAAAAAATTACCTTTTGAAGTAGTGAATTAAGAGAAAAAGCTTTAGTCCTCAGGT 1525
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 369 AAACAAAAAGGTTTACCTTATGAGACAGTGAATTAAGAGAGAAATGTTTAGTACCTCAGGT 310

QY 1526 TGTCAACAGCTAGAGAAAAAGTCTCAGTTTCAACAAATCCACATTCAAATGAGTTTTAAAT 1585
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 309 TGTCAACAGCTAGAGAAAGTCTCAGGCGCCACCAAAAT--ACATTTCAAGAAGAGTTTTAAAT 252

QY 1586 TTAATTTGTAATAAACTCATATTTACTGCCAAATATAGAAGAAAAATATTTTAAGTATTTGGTTA 1645
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 TTAATTTGTAATAAATGAATTTACTGCCAAATGTTAGAGAAAAACATCTTAGTATTTGGTTA 192

QY 1646 TGTGTAAATTTTCAATGCTGAAATGCTAAATAGATAGGTTCATATATAT- ---TCAAATTC 1701
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 TGTGTAAATTTTCAATGCTGAATCTCAATTAGATTAGGTTCATATATATATACTCAATTC 132

QY 1702 TTCAATTAATGATATTTGTTGATGGCAGTTTGTT- -AAAGTACTATATCATGTGTATAT 1759
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
QY 1175 GATGCTTTTATTGATTTTACATTTCTCTCTGCTGGACACCAATTTCTGTTTAAA 1234
|||||
Db 153 AATGTCGTTATTGATTTTACATTTCTCTCTGCTGGACACCAATTTCTGTTTAAA 94
|||||
QY 1235 TACCACCATTTTATGTTTAGGCCCAAGTGACCTTTTAGTAAATTAAGATGTGTTTTTT 1294
|||||
Db 93 TACCACCATTTTATGTTTAGGCCCAAGTGACCTTTTAGTAAATTAAGATGTGTTTTTCT 34
|||||
QY 1295 AGTCATGGCTTTATGGAAACAATATATTTTCACCC 1327
|||||
Db 33 AGTCATGGCTTTATGGAAACAATATATTTTCACCC 1

RESULT 9
LOCUS BG086308
DEFINITION H3124D01-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
ACCESSION BG086308
VERSION BG086308.1 GI:12568872
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
,T.S., Carter,M.G. and Ko,M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
Other_ESTs: H3124D01-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
plate: H3124 row: D column: 01
Seq primer: -21M13 Reverse
High quality sequence stop: 519
POLYA-No.

FEATURES
source
Location/Qualifiers
1..519
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:H3124D01-5"
/db_xref="taxon:10090"
/clone="H3124D01"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA
libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5 embryos
, and E12.5 female mesonephros/gonad) and one newborn
ovary cDNA library. Average insert size 1.5 kb. All
source libraries are cloned unidirectionally with Oligo(df
)-Not primers. References include: (1) Genome-wide
expression profiling of mid-gestation placenta and embryo
using a 15,000 mouse developmental cDNA microarray, 2000,
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
Large-scale cDNA analysis reveals phased gene expression
patterns during preimplantation mouse development, 2000,
Development, 127: 1737-1749; (3) Genome-wide mapping of
unselected transcripts from extraembryonic tissue of
7.5-day mouse embryos reveals enrichment in the t-complex
and under-representation on the X chromosome, 1998, Hum
```

```
BASE COUNT 135 a 120 c 93 g 171 t
ORIGIN
Query Match 24.0%; Score 451.8; DB 12; Length 519;
Best Local Similarity 91.9%; Pred. No. 1.1e-52;
Matches 477; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 288 ACAGTCGGAGATGACATTTGATGACATCAACACCAATATGTACCACCACTATCATATCCG 347
|||||
Db 1 ACGGTGCGAGATGACATTTGATGACATCGACCAATATGTACCACCACTATCATATCCCA 60
|||||
QY 348 TTAAGCTTTTCAAGTGTCTCTCACCGGATTTCTATCTTAGAAATTTCTGTTGGGACTTGC 407
|||||
Db 61 CTAAGCTTTTCAAGTGTCTCTCACTGGATTTCTCATGTTTAGAGATCGTCTGGGGCTTGGC 120
|||||
QY 408 AGCAACCTCAGTGTATGTGTACTTTACTGCATGAAATCCAACCTTAATCAACTCTGTCACT 467
|||||
Db 121 AGCAACCTTACCGTTCTGGTACTTTACTGCATGAAATCCAACCTTAATCAACTCTGTCACT 180
|||||
QY 468 AACATTATTACAATGAATCTTCATGTACTTGTATGATTAATAATTTGTGGGATGATTCCT 527
|||||
Db 181 AACATTATTACAATGAACCTCCATGTACTTGTATGATTAATAATTTGTGGGATGATTCCT 240
|||||
QY 528 CTAATATATAGTATCTCTCTGCTTTTCACTGGAGAGTAACACACTGCTCTCATTTGCTGTTTC 587
|||||
Db 241 CTAATATATAGTATCTCTCTGCTTTTCACTGGAGAGTAACACACTGCTCTCATTTGCTGTTTC 300
|||||
QY 588 CATGAGCTTGTGTATCTTTTGCAGTGTCTCAACAGCAATCAACGTTTTTGTGCTATFACAT 647
|||||
Db 301 CACGAAGCTTGTGTTTCCCTTTCAGTGTCTGACAGCAATCAACGTTTTTGTGCTATFACAT 360
|||||
QY 648 TTGGACAGATATGACATCTCTGTAACACCTGCAACCCGAAATCTTGACATGGGACAGCT 707
|||||
Db 361 CTGGACAGATATGACATCTCTGTAACACCTGCAACCCGAAATCTTGACATGGGACAGCT 420
|||||
QY 708 GTAATGTTAATGATATCCATTTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 767
|||||
Db 421 GTAATGTTAATGATATCCATTTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 480
|||||
QY 768 GAGGTAAATTTTTCAGTCTTCAAAAGTGGAAATACCTGG 806
|||||
Db 481 GAAGTAAATTTTTCAGTCTTCAAAAGTGGAAATACATGG 519
|||||

RESULT 10
BE824614/c
LOCUS BE824614
DEFINITION BE824614-3 NIA Mouse E13.5 VMB Dopamine cell cDNA Library Mus
musculus cDNA clone C0502B03 3', mRNA sequence.
ACCESSION BE824614
VERSION BE824614.1 GI:10256848
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 587)
Barrett,T., Xie,T., Piao,Y., Dillon-Carter,O., Kargul,G.J., Lim
,M.K., Chrest,F.J., Westo,R., Rowley,D.L., Juhaszova,M., Zhou,L.,
Vavter,M.P., Becker,K.G., Cheadle,C., Wood,W.H. III, McCann,U.D.,
Freed,W.J., Ko,M.S.H., Ricaurte,G. and Donovan,D.M.
Genes expressed by midbrain dopamine neurons: a dopamine
neuron-specific cDNA library and microarray used to examine changes
in gene expression during development and methamphetamine
neurotoxicity
Unpublished (2000)
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)
```

Plate: C0502 row: B column: 03
Seq primer: -21M13 Forward
High quality sequence stop: 587
POLYA=Yes.

FEATURES

source Location/Qualifiers

1..587

/organism="Mus musculus"

/strain="TH-beta-gal transgenic mouse"

/db_xref="niaEST:C0502B03-3"

/db_xref="taxon:10090"

/clone_lib="C0502B03"

/clone_lib="NIA Mouse E13.5 VMB Dopamine cell cDNA Library"

/sex="Male/Female"

/dev_stage="13.5dpc"

/lab_host="DHI0B"

/note="Vector: pSPORT1 (Gibco/BRL Life Technology); Site.1: SalI; Site.2: NotI; Total RNAs were extracted from 3000 Dopamine cells (cell collected by Dr.Tanya Barrett). The double-stranded cDNA was synthesized by Gibco's kit with an oligo(dT) primer [NotI primer-adaptor from GibcoBRL]

[5'-pgactagttctagatcgagcgccgcttttttttttt-3'] from 0.91ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-Sal4 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DHI0B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Yulan Piao and Minoru.S. H. Ko."

BASE COUNT 192 a 101 c 85 g 209 t

ORIGIN

Query Match 23.2%; Score 435.8; DB 12; Length 587;
Best Local Similarity 87.3%; Pred. No. 1.5e-50;
Matches 514; Conservative 0; Mismatches 67; Indels 8; Gaps 3;

QY 1248 TGTTCAGGCGCAAGTACCTTTAGTAAATTAAGATGTGTTTTAGTCATGGCTTAT 1307

Db TGTTCAGGCGCAAGTACCTTTAGTAAATTAAGATGTGTTTTAGTCATGGCTTAT 528

QY 1308 GGAACAACTATATTTACCCCTCTATTATATGCATTCCTAGACAGAAAATTTCAAAGGTC 1367

Db TGTTCAGGCGCAAGTACCTTTAGTAAATTAAGATGTGTTTTAGTCATGGCTTAT 528

QY 1368 TCGAAGGTAATTAAGAAAGCGAGTGTGTTCTATAGTAGAAGCTGATCCCTGCCTTAAT 1427

Db TTAAGAGTAGAATGAAGAAAGCGAGTGTGTTCTATAGTAGAAGCTGATCCCTGCCTTAAT 408

QY 1428 AATGCTGTATACAACTCTTGGATAGATCCCAAGAAACAAAATTAACCTTTGAA 1487

Db AACGCTGTATACAACTCATGGATAGATCCCTAAAGAAACAAAAGGTTACCTATGAA 348

QY 1488 GATAGTGAATATAGAGAAAACGTTTAGTGCCTTCAGGTTGTCAGACTAGAGAAAGTC 1547

Db GACAGTGAATATAGAGAAAATGTTTAGTACCTCAGGTTGTCAGACTAGAGAAAGTC 288

QY 1548 TCAGTTTCACCAATCCACATTCAGTTTAAATTTAAATTTAAATTTGTAACCTGATATT 1607

Db TCAGGCGCCCAAT--ACATTCAGAGAGAGTTTAAATTTAAATTTGTAACCTGATATT 230

QY 1608 ACTGCCAAATATAGAAAATATTTTAAGTATTTGGTTATGTTGTAATTTCAATGTGAA 1667

Db CTGCCAAATCTTAGAGAAAACATCTTAGGTATTTGGTTATGTTGTAATTTCAATGTGAA 170

QY 1668 ATGCTAATTAGATAGTGCATATATAT---TCAATTTCTTCATTACTACTATGTTTGT 1723

Db TGCAATATTAGTATAGTGCATATATATATCAATTTCTTCATTACTATGTTTGT 110

QY 1724 GCATGGCAGTTTGT--AAAGTACTATCATGTGTATATATTTTGTCAATATTATGTCCAACA 1781

Db GCATGGCAGTTTGTAAAGAAAGTAATCATGTGTATATATTTTGTCAATATTATGTCCAACA 50

QY 1782 GAAATATTCATGTAGTCAATATTTTAAAGGAATAAATACATAGCCTT 1830

Db GAAGATATCCATGTAGTCAATATTTTCTTAAAGAAATAAATATGTAGCCTT 1

RESULT 11

BE824967/c

LOCUS

DEFINITION

BE824967

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE824967 591 bp mRNA linear EST 21-SEP-2000

C0507E08-3 NIA Mouse E13.5 VMB Dopamine cell cDNA Library Mus

musculus cDNA clone C0507E08 3', mRNA sequence.

BE824967

BE824967.1 GI:10257201

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 591)

Barrett,T., Xie,T., Piao,Y., Dillon-Carter,O., Kargul,G.J., Lim ,M.K., Christ,F.J., Wersto,R., Rowley,D.L., Juhaszova,M., Zhou,L., Vawter,M.P., Becker,K.G., Cheadle,C., Wood,W.H. III, McCann,U.D., Freed,W.J., Ko,M.S.H., Ricaurte,G. and Donovan,D.M.

Genes expressed by midbrain dopamine neurons: a dopamine neuron-specific cDNA library and microarray used to examine changes in gene expression during development and methamphetamine neurotoxicity

Unpublished (2000)

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Casseil Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov

niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)

Plate: C0507 row: E column: 08

Seq primer: -21M13 Forward

High quality sequence stop: 591

POLYA=Yes.

FEATURES

source Location/Qualifiers

1..591

/organism="Mus musculus"

/strain="TH-beta-gal transgenic mouse"

/db_xref="niaEST:C0507E08-3"

/db_xref="taxon:10090"

/clone_lib="C0507E08"

/clone_lib="NIA Mouse E13.5 VMB Dopamine cell cDNA Library"

/sex="Male/Female"

/dev_stage="13.5dpc"

/lab_host="DHI0B"

/note="Vector: pSPORT1 (Gibco/BRL Life Technology); Site.1: SalI; Site.2: NotI; Total RNAs were extracted from 3000 Dopamine cells (cell collected by Dr.Tanya Barrett). The double-stranded cDNA was synthesized by Gibco's kit with an oligo(dT) primer [NotI primer-adaptor from GibcoBRL]

[5'-pgactagttctagatcgagcgccgcttttttttttt-3'] from 0.91ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-Sal4 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were

digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Yulan Piao and Minoru.S. H. Ko."

BASE COUNT	194 a	101 c	86 g	210 t	
ORIGIN					
Query Match	23.1%; Score 435; DB 12; Length 591;				
Best Local Similarity	86.8%; Pred. No. 2e-50;				
Matches	515;	Conservative	0;	Mismatches	70; Indels 8; Gaps 3;
QY	1244	TTATGTTTGGCCCAAGTGACCTTTT	TAGTAAATTAAGATGTGTTTTT	TAGTCATGCG	1303
Db	591	TCATGTTTGGCCCAAGTGACCTTTT	TAGTAAATTAAGATGTGTTTTT	TAGTCATGCG	532
QY	1304	TTATGGACAACTATATTTCCACCTCT	TATATATATGATCAGTACAGAAATTT	CAAAA	1363
Db	531	TTATGGACAACTATATTTCCACCTCT	TATATATATGATCAGTACAGAAATTT	CAAAA	472
QY	1364	GGTCCTTGAAAGTAAATGAAAGCGAG	TGTTTCTATAGTAGAGCTGATCCCTGCG	1423	
Db	471	GGTCCTTGAAAGTAAATGAAAGCGAG	TGTTTCTATAGTAGAGCTGATCCCTGCG	412	
QY	1424	TAATTAATGCTGTAATACAACTCTT	TGGATAGATCCCAAGAAACAAAAATTT	ACCTT	1483
Db	411	TAATTAATGCTGTAATACAACTCTT	TGGATAGATCCCAAGAAACAAAAATTT	ACCTT	352
QY	1484	TGAAGATAGTGAATAAGAGAAACGTT	TAGTGCTCAGGTGTCAGAGCTAGAGAA	1543	
Db	351	TGAAGATAGTGAATAAGAGAAACGTT	TAGTGCTCAGGTGTCAGAGCTAGAGAA	292	
QY	1544	AGTCCTCAGTTTCACCAATCCACATTC	CAAAATGAGTTTAAATTTAAATTTGAA	1603	
Db	291	AGTCCTCAGTTTCACCAATCCACATTC	CAAAATGAGTTTAAATTTAAATTTGAA	234	
QY	1604	TATTACTGCCAATATAGAAATATTTT	TAGTATGTTGTTGTAATTTTCAATG	1663	
Db	233	ATTACTGCCAATATAGAAATATTTT	TAGTATGTTGTTGTAATTTTCAATG	174	
QY	1664	TGAATGCTATATAGATAGGTCATATAT	ATATATTTTCAATTTTCAATTTTCAAT	1719	
Db	173	TGAATGCTATATAGATAGGTCATATAT	ATATATTTTCAATTTTCAATTTTCAAT	114	
QY	1720	TGTTGCATGCGCAGTTTGT--AAAGT	ACTATCATGTGTATATTTTGTCAATAT	1777	
Db	113	TGTTGCATGCGCAGTTTGT--AAAGT	ACTATCATGTGTATATTTTGTCAATAT	54	
QY	1778	ACAGAAATATTCATGTAAGTCATATTT	TTTAAAGGAATAATACATAGCCTT	1830	
Db	53	ATCAGAAGATATCCATGTAAGTCATAT	TTTCTTAAAGGAATAATATGAGCCTT	1	

RESULT 12
BE824766/c
LOCUS BE824766 587 bp mRNA linear EST 21-SEP-2000
DEFINITION C0504E03-3 NIA Mouse E13.5 VMB Dopamine cell cDNA Library Mus
ACCESSION BE824766
VERSION BE824766.1 GI:10257000
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 587)
Barrett,T., Xie,T., Piao,Y., Dillon-Carter,O., Kargul,G.J., Lim
M.K., Chrest,F.J., Wersto,D.L., Juhaszova,M., Zhou,L.,
Vawter,M.P., Becker,K.G., Cheadle,C., Wood,W.H. III, McCann,U.D.,
Freed,W.J., Ko,M.S.H., Ricaurte,G. and Donovan,D.M.
Genes expressed by midbrain dopamine neurons: a dopamine
neuron-specific cDNA library and microarray used to examine changes

JOURNAL
COMMENT

in gene expression during development and methamphetamine neurotoxicity
Unpublished (2000)
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)
Plate: C0504 row: E column: 03
Seq primer: -21M13 Forward
High quality sequence stop: 587
POLYA-Yes.

FEATURES
Source

Location/Qualifiers
1. 587
/organism="Mus musculus"
/strain="TH-beta-gal transgenic mouse"
/db_xref="niaEST:C0504E03-3"
/db_xref="taxon:10090"
/clone="C0504E03"
/clone.lib="NIA Mouse E13.5 VMB Dopamine cell cDNA Library"
/sex="Male/Female"
/dev_stage="13.5dpc"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Gibco/BRL Life Technology); Site 1: SalI; Site 2: NotI; Total RNAs were extracted from 3000 Dopamine cells (cell collected by Dr.Yanya Barrett). The double-stranded cDNA was synthesized by Gibco's kit with an Oligo(dT) primer [NotI primer-adaptor from GibcoBRL]
[5'-'PGAGTAGTTCTAGATCGAGCGGCCCTTTTTTTTTTTT-3'] from 0.91ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker Li-Sal4 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Yulan Piao and Minoru.S. H. Ko."

BASE COUNT
ORIGIN

Query Match	22.9%;	Score 431.6;	DB 12;	Length 587;	
Best Local Similarity	86.8%;	Pred. No. 5.7e-50;			
Matches	511;	Conservative	0;	Mismatches	70; Indels 8; Gaps 3;
QY	1248	TGTTTAGGCCCAAGTGACCTTTT	TAGTAAATTAAGATGTGTTTTT	TAGTCATGCTTAT	1307
Db	587	TGTTTAGGCCCAAGTGACCTTTT	TAGTAAATTAAGATGTGTTTTT	TAGTCATGCTTAT	528
QY	1308	GGAACAACATATATTTT	CACCTCTATATATGATTCATC	ACTAGACAAAAATTTCAAAGGTC	1367
Db	527	GGAACAACATATATTTT	CACCTCTATATGATTCATC	ACTAGACAAAAATTTCAAAGGTC	468
QY	1368	TGAAAAGTAAATGAAAAGCGAG	TGTTTCTATAGTAGAAGCTGAT	CCCTCGCTTAAT	1427
Db	467	TGAAAAGTAAATGAAAAGCGAG	TGTTTCTATAGTAGAAGCTGAT	CCCTCGCTTAAT	408
QY	1428	AATGCTGTAATACAACTCTT	TGGATAGATCCCAAGAAACAAAA	AAATTTACCTTTGAA	1487
Db	407	AATGCTGTAATACAACTCTT	TGGATAGATCCCAAGAAACAAAA	AAATTTACCTTTGAA	348
QY	1488	GATAGTGAATAAGAGAAACG	TTTAGTCTCAGGTGTCACAGACT	AGAGAAAGTC	1547
Db	347	GATAGTGAATAAGAGAAACG	TTTAGTCTCAGGTGTCACAGACT	AGAGAAAGTC	288


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p313j1132"
/clone_lib="313 (synonym: hlcc2)"
/dev_stage="adult"
/lab_host="DH108"
/notes="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI;
cDNA-collection"
BASE COUNT 134 a 54 c 46 g 188 t
ORIGIN
Query Match 20.1%; Score 379; DB 9; Length 422;
Best Local Similarity 95.0%; Pred. No. 1.1e-42;
Matches 402; Conservative 0; Mismatches 20; Indels 1; Gaps 1;
QY 1454 AGATCCAAAGAAACAAAAAATACCTTTGAAGATAGTGAATAGAGAAAAAGCTTT 1513
DB 422 AGATCCTAAAGAAACAAAAAATACCTTTGAAGATAGTGAATAGAGAAAAAGCTTT 363
QY 1514 AGTGCTCAGGTGTGCACAGACTAGAGAAAGTCTCAGTTTCCACAAATCCACATCAAA 1573
DB 362 AGTGCTCAGGTGTGCACAGACTAGAGAAAGTCTCAGTTTCCACAAATCCACATCAAA 303
QY 1574 TGAGTTTAAATTTAAATGTTAAACTCATATTACTGCGCAATATAGAAAAATATTT 1633
DB 302 TGAGTTTAAATTTAAATGTTAAACTCATATTACTGCGCAATATAGAAAAATATTT 243
QY 1634 AAGTATTGGTTATGTTGTAATTTCAATGTGAATGCTAATAGATAGTGCATATATAT 1693
DB 242 AAGTATTGGTTATGTTGTAATTTCAATGTGCAATAGTGCATATATATATATAT 184
QY 1694 TCAATTTCTTCAATTAATGATTTGTTGTCATGCGAGTGTGTTAAAGTACTATCATGT 1753
DB 183 TCAATTTCTTCAATTAATGATTTGTTGTCATGCGAGTGTGTTAAAGTACTATCATGT 124
QY 1754 GTATATTTGTCATATTTATGTCACACAAAAATATTCATGTAAGTCATATTTTAAAG 1813
DB 123 GTATATTTGTCATATTTATGTCACACAAAAATATTCATGTAAGTCATATTTTAAAG 64
QY 1814 AATAATACATACGCTTAAACAGTGTATAACTTTTAAATGTAAAAATGTAACAAAAA 1873
DB 63 AATAATACATACGCTTAAACAGTGTATAACTTTTAAATGTAAAAATGTAACAAAAA 4
QY 1874 AAA 1876
DB 3 ACA 1
RESULT 15
AW061316/c
LOCUS
DEFINITION
UI-M-BH1-anw-d-12-0-UI-s1 NIH_BMAP_M_S2 Mus musculus cDNA clone
UI-M-BH1-anw-d-12-0-UI 3', mRNA sequence.
ACCESSION
AW061316
VERSION
AW061316.1 GI:6009067
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 520)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
```

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized hippocampus library cDNA library preparation: M.B. Soares lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. Seq primer: M13 Forward POLYA=yes.

FEATURES
source

Location/Qualifiers
1..520

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH1-anw-d-12-0-UI"
/clone_lib="NIH_BMAP_M_S2"
/dev_stage="27-32 days"
/lab_host="DH108 (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S2 library is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH_BMAP_M_S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries.
TAG_LIB=NIH_BMAP_M_S2
TAG_TISSUE=hippocampus
TAG_SEQ=TTCCA"

BASE COUNT 162 a 86 c 68 g 204 t
ORIGIN

Query Match 19.3%; Score 362.6; DB 10; Length 520;
Best Local Similarity 85.9%; Pred. No. 1.6e-40;
Matches 439; Conservative 0; Mismatches 64; Indels 8; Gaps 3;

QY 1332 TTATATGCAATTCACATAGACAAAAATTTCAAAGGTCTTGAAAAAGTAAATGAAAAAGCGA 1391
DB 520 TTGTATGCAATTCACACAGACAAAAGTTTCAAAGGTCTTAAAGGTAGATGAAAAAGCGA 461
QY 1392 GTTGTCTTATAGTAGAAGCTGATCCCTGCGCTTAATGCTGTATACACAACTCTTG 1451
DB 460 GTTGTCTTATAGTAGAAGCTGATCCCTGCGCTTAATGCTGTATACACAACTCATGG 401
QY 1452 ATAGATCCCAAGAACAAAAAATTTACCTTTGAAGATAGTGAATAGAGAAAAAGCT 1511
DB 400 ATAGATCCCAAGAACAAAAAAGGTTCACCTAGAGACAGTGAATAGAGAAAAAGT 341
QY 1512 TTAGTGCCTCAGGTGTGCACAGACTAGAAAAAGTCTCAGTTTCCACAAATCCACATCA 1571
DB 340 TTAGTGCCTCAGGTGTGCACAGACTAGAGAGAGTCTCAGCCCACTCAAT--ACATTTCA 283
QY 1572 AATGAGTTTTAAATTTAAATTTGTAATAATGATATTACTGCGCAAAATATAGAAAAATAT 1631
DB 282 GAAGAGTTTTAAATTTAAATTTGTAATAATGATATTACTGCGCAAAATGTTAGAGAAAAAT 223
QY 1632 TTAAGTATTTGGTATGTTCTTAATTTCAATGTAATGTAATGTAATGTAATGTAATGTAAT 1691
DB 222 CTAGGATTTGGTATGTTCTTAATTTCAATGTAATGTAATGTAATGTAATGTAATGTAAT 163
QY 1692 AT----TCAATTTCTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAAT 1745
DB 162 ATATACTCAATTTCTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTCAAT 103
QY 1746 TATCATGTGTATATTTTGTCAATATTTATTCACAAACAAAAATATTCATGTAAGTCAATAT 1805

Db 102 TATCATGTGTATATTTTGTCAATATTATGTCATCAGAGATATCCATGTAAGTCATATT 43

Qy 1806 TTTTAAGGAATAATACATAGCCTTAAACA 1836

Db 42 TTCTAAAGAATAAATATGTAGCCTTAAAAA 12

Search completed: December 10, 2002, 09:44:13
Job time : 2418 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 10, 2002, 07:34:42 ; Search time 4717 Seconds
(without alignments)
11605.338 Million cell updates/sec

Title: US-09-845-721-1
Perfect score: 1881
Sequence: 1 gttattttcctcaaaagaaa.....aaaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.mu.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1881	100.0	1881	6	AX298123 Sequence
2	1881	100.0	1881	9	HSU66581 Human putat
3	1826.6	97.1	172533	9	AC002381 Human BAC
4	1625	86.4	2238	6	AR091359 Sequence
5	1388.8	73.8	83627	2	AC096185 Rattus no
6	535.8	28.5	76186	2	AC125020 Mus muscu
c 7	78.8	4.2	97683	2	AC116548 Dictyoste
c 8	76	4.0	201299	9	AC007158 Homo sapi
c 9	75.6	4.0	136098	9	AC006970 Homo sapi
c 10	75	4.0	93491	2	AC116967 Dictyoste
c 11	74.4	4.0	34980	6	AX344555 Sequence
c 12	73.2	3.9	106958	9	AL807813 Human DNA
c 13	72	3.9	201299	9	AC007158 Homo sapi
c 14	72	3.8	91723	9	AC010903 Homo sapi
c 15	71.8	3.8	25117	2	AC116102 Dictyoste
c 16	71.6	3.8	198431	9	AC109994 Homo sapi
c 17	71	3.8	11829	3	AE001425 Plasmodiu
c 18	71	3.8	191366	2	AC115351 Rattus no
c 19	70.6	3.8	2982	3	DDISGSPA M33862 Dictyosteli
c 20	70.4	3.7	141017	2	AC116962 Dictyoste
c 21	70.2	3.7	84805	2	AC116918 Dictyoste
c 22	70.2	3.7	15317	2	AC067973 Homo sapi
c 23	69	3.7	6007	3	AF474335 Dictyoste
c 24	69	3.7	118001	9	AC073081 Homo sapi
c 25	68.6	3.6	123280	2	AC117076 Dictyoste
c 26	68.6	3.6	206059	2	AC127383 Homo sapi
c 27	68.6	3.6	318221	2	PFMAL13P3 AL049184 Plasmodiu
c 28	68.4	3.6	149752	9	AC004616 Homo sapi
c 29	68.2	3.6	163443	2	AC006280 Plasmodiu
c 30	68.2	3.6	349980	6	AX344573 Sequence
c 31	68	3.6	133148	9	AC009435 Homo sapi
c 32	67.6	3.6	5487	6	AX346500 Sequence
c 33	67.6	3.6	176898	9	AC117569 Homo sapi
c 34	67.4	3.6	96183	9	AL161896 Human DNA
c 35	67.4	3.6	130349	9	AC011593 Homo sapi
c 36	67.2	3.6	131682	9	AL672277 Human DNA
c 37	67.2	3.6	198431	9	AC109994 Homo sapi
c 38	67.2	3.6	321003	2	PFMAL4P3 AL035476 Plasmodiu
c 39	67	3.6	2452	3	AF337815 Dictyoste
c 40	66.8	3.6	6047	2	AC115582 Dictyoste
c 41	66.8	3.6	40033	9	HS1110P6 AL049175 Human DNA
c 42	66.8	3.6	159942	9	AC025018 Homo sapi
c 43	66.8	3.6	187190	2	AC026898 Homo sapi
c 44	66.6	3.5	12029	3	AE001379 Plasmodiu
c 45	66.6	3.5	72243	9	AL731858 Human DNA

ALIGNMENTS

RESULT 1	AX298123	Sequence 1 from Patent WO0183550.	1881 bp	DNA	linear	PAT 26-NOV-2001
LOCUS	AX298123	Sequence 1 from Patent WO0183550.				
DEFINITION	AX298123					
ACCESSION	AX298123.1	GI:17128199				
VERSION	AX298123.1					
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1					
AUTHORS	Brennan, J.C. and Hart, K.A.					
TITLE	Gpr22 modulators as appetite-control agents					
JOURNAL	Patent: WO 0183550-A 1 08-NOV-2001;					

FEATURES	Astrazeneca AB (SE)									
source	location/Qualifiers									
	1. .1881									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
BASE COUNT	649 a	317 c	288 g	627 t						
ORIGIN										
Query Match	100.0%; Score 1881; DB 6; Length 1881;									
Best Local Similarity	100.0%; Pred. No. 1.3e-304;									
Matches 1881;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;		
QY	1	GTATTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG	60							
Db	1	GTATTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG	60							
QY	61	GTCTTAATTTCTACATTTTCTATTAAATAGTTTACAAACTTAAAAATTAAGTACATAC	120							
Db	61	GTCTTAATTTCTACATTTTCTATTAAATAGTTTACAAACTTAAAAATTAAGTACATAC	120							
QY	121	AATTGAAGAATTTTCTTTTACAAAGAACACAGTTTATAGCTATTTAAATTTGCCAAATA	180							
Db	121	AATTGAAGAATTTTCTTTTACAAAGAACACAGTTTATAGCTATTTAAATTTGCCAAATA	180							
QY	181	TCAAATAGTTTATTTCTATTTCACHTTCTAGGGAAGAAACCAACTGCTCCAAAGAAATGT	240							
Db	181	TCAAATAGTTTATTTCTATTTCACHTTCTAGGGAAGAAACCAACTGCTCCAAAGAAATGT	240							
QY	241	GTCTTTCTCCCATTCGGAATCAACATGCAGTCTGAATCTAACATTCACAGTCGGAGATG	300							
Db	241	GTCTTTCTCCCATTCGGAATCAACATGCAGTCTGAATCTAACATTCACAGTCGGAGATG	300							
QY	301	ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTTCAAG	360							
Db	301	ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTTCAAG	360							
QY	361	TGCTCTCACGGGATTTCTTATGTTAGAAATTTGTGTTGGGACTTGGCAGCAACCTCACTG	420							
Db	361	TGCTCTCACGGGATTTCTTATGTTAGAAATTTGTGTTGGGACTTGGCAGCAACCTCACTG	420							
QY	421	TATTTGGTACTTTTACTGCATGAATCCAACTTAATCAACTCTGTCACTAATATTATACAA	480							
Db	421	TATTTGGTACTTTTACTGCATGAATCCAACTTAATCAACTCTGTCACTAATATTATACAA	480							
QY	481	TGAATCTTCATGTACTTTGATGTAATAATTTGTGTGGGATGTATTCCCTCAACTATAGTTA	540							
Db	481	TGAATCTTCATGTACTTTGATGTAATAATTTGTGTGGGATGTATTCCCTCAACTATAGTTA	540							
QY	541	TCCTTCTGCTTTCACTGGAGAGTAACACTGCTCTCACTTTGCTGTTCCATGAGGCTTTGTG	600							
Db	541	TCCTTCTGCTTTCACTGGAGAGTAACACTGCTCTCACTTTGCTGTTCCATGAGGCTTTGTG	600							
QY	601	TATCTTTTTCGAAGTGTCTCAACAGCAATCAACGTTTGTCTATCACTTTTGGACAGATATG	660							
Db	601	TATCTTTTTCGAAGTGTCTCAACAGCAATCAACGTTTGTCTATCACTTTTGGACAGATATG	660							
QY	661	ACATCTCTGTAAACCTGCAAAACCGGAATTTCTGACAAATGGGACAGCTGTAAATGTAATGA	720							
Db	661	ACATCTCTGTAAACCTGCAAAACCGGAATTTCTGACAAATGGGACAGCTGTAAATGTAATGA	720							
QY	721	TATCCATTTGGATTTTCTTTTCTCTTCTCTGATTCCTTTTATGAGGTAATTTT	780							
Db	721	TATCCATTTGGATTTTCTTTTCTCTTCTCTGATTCCTTTTATGAGGTAATTTT	780							
QY	781	TCAGTCTTCAAGTGGAAATACCTGGGAAACAAAGACACTTTTATGTCAGTACAAATG	840							
Db	781	TCAGTCTTCAAGTGGAAATACCTGGGAAACAAAGACACTTTTATGTCAGTACAAATG	840							
QY	841	AATACACTGAACCTGGGAATGTATTATCACTGTTTAGTACAGATCCCAATATTTCTTTT	900							
Db	841	AATACACTGAACCTGGGAATGTATTATCACTGTTTAGTACAGATCCCAATATTTCTTTT	900							
QY	901	TCAGTGTGTAGTAATGTTTAAACACATACACCAAAATACCTTCAGGCTCTTAATATTCGAA	960							

RESULT 2

HSU66581

LOCUS

DEFINITION

Human putative G protein-coupled receptor (GPR22) gene, complete cds.

1881 bp DNA linear PRI 15-MAY-1997

Db	901	TCACCTGTTAGTAATGTTAATCACATACACCAAAATACTTCAGGCTCTTAATATTCGAA	960
QY	961	TAGGCACAAGATTTTCAACAGGCGAGAAAGCAAGCAAGAAAGAAAGCAATTTCTC	1020
Db	961	TAGGCACAAGATTTTCAACAGGCGAGAAAGCAAGCAAGAAAGCAATTTCTC	1020
QY	1021	TAAACACACAATGAGGCTACAGACATGTCACAAAGCAGTGGTGGAGAAATGTAGTCT	1080
Db	1021	TAAACACACAATGAGGCTACAGACATGTCACAAAGCAGTGGTGGAGAAATGTAGTCT	1080
QY	1081	TTGGTGTAAACAATTCAGTTTCTGTAATATTTGCCCTCCGCGAGCTGTGAACGACACC	1140
Db	1081	TTGGTGTAAACAATTCAGTTTCTGTAATATTTGCCCTCCGCGAGCTGTGAACGACACC	1140
QY	1141	GTCAACGACGAGAAAGACAAAGAGAGCTTTCAGGATGCTTTTATGATTTCTACAT	1200
Db	1141	GTCAACGACGAGAAAGACAAAGAGAGCTTTCAGGATGCTTTTATGATTTCTACAT	1200
QY	1201	TTCTTCTCTGCTGGACACCAATTTCTGTTTAAATACCACTTTTATGTTAGGCCCAA	1260
Db	1201	TTCTTCTCTGCTGGACACCAATTTCTGTTTAAATACCACTTTTATGTTAGGCCCAA	1260
QY	1261	GTGACCTTTTAGTAAATTAAGATTTGTTTATGTCATGCGCTTATGGAACACATATAT	1320
Db	1261	GTGACCTTTTAGTAAATTAAGATTTGTTTATGTCATGCGCTTATGGAACACATATAT	1320
QY	1321	TTCAACCTCTATTTATGCAATTCACAGACAAATTTTCAAAAGGCTTTCAAAAGTAAAA	1380
Db	1321	TTCAACCTCTATTTATGCAATTCACAGACAAATTTTCAAAAGGCTTTCAAAAGTAAAA	1380
QY	1381	TGAAAAGCGAGTTGTTTCTATAGTAAAGCTGATCCCTCGCTTAATATGCTGTATATC	1440
Db	1381	TGAAAAGCGAGTTGTTTCTATAGTAAAGCTGATCCCTCGCTTAATATGCTGTATATC	1440
QY	1441	ACAACTCTTGGATAGATCCCAAGAAACAAATAATTTACCTTTGAAGATAGTGAATAA	1500
Db	1441	ACAACTCTTGGATAGATCCCAAGAAACAAATAATTTACCTTTGAAGATAGTGAATAA	1500
QY	1501	GAGAAAACCTTTAGTGCCTCAGGTTGTACACACTAGAGAAAAGTCTCAGTTTCACCAA	1560
Db	1501	GAGAAAACCTTTAGTGCCTCAGGTTGTACACACTAGAGAAAAGTCTCAGTTTCACCAA	1560
QY	1561	ATCCACATTCAAATGAGTTTAAATTTAAATTTGTAATAATGATATGCTGCAAAATATA	1620
Db	1561	ATCCACATTCAAATGAGTTTAAATTTAAATTTGTAATAATGATATGCTGCAAAATATA	1620
QY	1621	AGAAAATATTTTAAAGTATGTTTGTGTAATTTTCAATGTGAAATGCTAATTAGAT	1680
Db	1621	AGAAAATATTTTAAAGTATGTTTGTGTAATTTTCAATGTGAAATGCTAATTAGAT	1680
QY	1681	AGGTCAATATATTTCAATTTCTTCAATTAATGTTTGTGTTGTTGTTGTTGTTGTTAA	1740
Db	1681	AGGTCAATATATTTCAATTTCTTCAATTAATGTTTGTGTTGTTGTTGTTGTTGTTAA	1740
QY	1741	AGTACTATCATGTGTATTTTGTCAATATATGTCACACAGAAAATATTCATGTAGTC	1800
Db	1741	AGTACTATCATGTGTATTTTGTCAATATATGTCACACAGAAAATATTCATGTAGTC	1800
QY	1801	ATATTTTAAAGAAATTAATATACAGCTTAAACAGTGTATAACTTTAAATGTAAAAA	1860
Db	1801	ATATTTTAAAGAAATTAATATACAGCTTAAACAGTGTATAACTTTAAATGTAAAAA	1860
QY	1861	AAAAAAAAAAAAAAAAAAAAA 1881	
Db	1861	AAAAAAAAAAAAAAAAAAAAA 1881	

```

ACCESSION   U66581
VERSION     U66581.1  GI:1753106
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 1881)
AUTHORS     O'Dowd,B.F., Nguyen,T., Jung,B.P., Marchese,A., Cheng,R.,
            Heng,H.H., Kolakowski,L.F. Jr., Lynch,K.R. and George,S.R.
TITLE       Cloning and chromosomal mapping of four putative novel human
            G-protein-coupled receptor genes
JOURNAL     Gene 187 (1), 75-81 (1997)
MEDLINE     97225799
PUBMED      9073069
REFERENCE   2 (bases 1 to 1881)
AUTHORS     O'Dowd,B.F., Nguyen,T., Jung,B., Marchese,A., Cheng,R.,
            Heng,H.H., Kolakowski,L.F. Jr., Lynch,K.R. and George,S.R.
TITLE       Direct Submission
JOURNAL     Submitted (12-AUG-1996) Department of Pharmacology, University of
            Toronto, 8 Taddle Creek Rd., Toronto, Ontario M5S 1A8, Canada
FEATURES
            source
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="7"
            /map="7q22-q31.1"
            /gene="GPR22"
            /gene="GPR22"
            /codon_start=1
            /product="putative G protein-coupled receptor"
            /protein_id="AAC51304.1"
            /db_xref="GI:1753107"
            /translation="MCFSPILNMQSESNITVRDDIDIDINMYQPLSPQVSL
            TGLFLEIVLGLGSLNVLVLCMKSNLNIITMNLHLDVILVCIGPIITVI
            LLLESENALICCFHEACVSFAISINVEAITLDYDLSVKPARIILTMGRVML
            MISIWFSPFLPIEFIEVNFSLQSNWENKILLYSVNEYTELGMTHLLVQIP
            IFFTVVMIITIKILQALNIRIGTFSGQKKARKKTIISLTQHEATDMSQSG
            GRNVFVRTSVSIVIALRAVRRRERQRKVRMSLLIISTFLLCWTPISVLT
            TIILGSLDILLRLICFLVMAYGTTIFHPLLYAFTQKFKQLKMKRVRVSIVEA
            DPLPNNAVINSWIDPKRKKITFEDSEIREKRLVPQVTD"
BASE COUNT  649 a 317 c 288 g 627 t
ORIGIN
Query Match 100.0%; Score 1881; DB 9; Length 1881;
Best Local Similarity 100.0%; Pred. No. 1.3e-304;
Matches 1881; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTTATTTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG 60
Db 1 GTTATTTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG 60
Qy 61 GTTCTTAATCTACATTTCTTATTAATAGTTTACAAACTTAAATTAACCTTAAGTACAC 120
Db 61 GTTCTTAATCTACATTTCTTATTAATAGTTTACAAACTTAAATTAACCTTAAGTACAC 120
Qy 121 AATTGAAAGATTTTTTTTCTTACAAAGAACACGTTATACGTCAITTTAAATTTGCCAAATA 180
Db 121 AATTGAAAGATTTTTTTTCTTACAAAGAACACGTTATACGTCAITTTAAATTTGCCAAATA 180
Qy 181 TCMAATAGTTTATCTATTCTACCTTCTAGGAGAAAACCAACTGCTCCAAAGAAATGT 240
Db 181 TCMAATAGTTTATCTATTCTACCTTCTAGGAGAAAACCAACTGCTCCAAAGAAATGT 240
Qy 241 GTTTTTCTCCCATCTCGAAATCAACATGCAGTCTGAATCTAAACATTACAGTGCAGATG 300
Db 241 GTTTTTCTCCCATCTCGAAATCAACATGCAGTCTGAATCTAAACATTACAGTGCAGATG 300
Qy 301 ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAG 360
Db 301 ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAG 360

```

361 TGCTCTCACCGGATTTCTTATAGAAATGTGTGGACTTGGCAGCAACCTCACTG 420
361 TGCTCTCACCGGATTTCTTATAGAAATGTGTGGACTTGGCAGCAACCTCACTG 420
421 TATTGGTACTTACTGTCATGAAATCCAATTAATCAACTCTCTCACTAGCAACATTAACAA 480
421 TATTGGTACTTACTGTCATGAAATCCAATTAATCAACTCTCTCACTAGCAACATTAACAA 480
481 TGAATCTTCATGTAATGATGTAATTTGTTGGGATGATATCTCTCACTAATAGTTA 540
481 TGAATCTTCATGTAATGATGTAATTTGTTGGGATGATATCTCTCACTAATAGTTA 540
541 TCCCTCTCTGTTTCACTGGAGAGTAACACTGCTCTCATTTGCTGTTTCCATGAGGCTTGG 600
541 TCCCTCTCTGTTTCACTGGAGAGTAACACTGCTCTCATTTGCTGTTTCCATGAGGCTTGG 600
601 TATCTTTTGGCAAGTGTCTCAACAGCAATCAACGTTTTTGTCTATCACTTTGGACAGATG 660
601 TATCTTTTGGCAAGTGTCTCAACAGCAATCAACGTTTTTGTCTATCACTTTGGACAGATG 660
661 ACATCTCTGTAACCACTGCAACCGAATTTCTGACAAATGGCAGAGCTGTAATGTTAATGA 720
661 ACATCTCTGTAACCACTGCAACCGAATTTCTGACAAATGGCAGAGCTGTAATGTTAATGA 720
721 TATCCATTTGGATTTTTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
721 TATCCATTTGGATTTTTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
781 TCAGTCTTCAAAAGTGGAAATACCTGGGAAACAAACACACTTTTATGTGTGTCAGTACAAATG 840
781 TCAGTCTTCAAAAGTGGAAATACCTGGGAAACAAACACACTTTTATGTGTGTCAGTACAAATG 840
841 AATACTACACTGAACCTGGGAATGATATACCTGTTAGTACAGATGCCAATATTTCTTTT 900
841 AATACTACACTGAACCTGGGAATGATATACCTGTTAGTACAGATGCCAATATTTCTTTT 900
901 TCAGTGTGTAGTAACTGTAATACATACACAAATACTTTCAGGCTCTTAATATTCGAA 960
901 TCAGTGTGTAGTAACTGTAATACATACACAAATACTTTCAGGCTCTTAATATTCGAA 960
961 TAGGCACAAAGATTTTCAACAGGGCAGAGAAAGAAAGCAAGAAAGAAAGCAATTTCTC 1020
961 TAGGCACAAAGATTTTCAACAGGGCAGAGAAAGAAAGCAAGAAAGAAAGCAATTTCTC 1020
1021 TAACACACAACTGAGGCTACAGACATGTACAAAGCAGTGGTGGGAGAAATGTAGTCT 1080
1021 TAACACACAACTGAGGCTACAGACATGTACAAAGCAGTGGTGGGAGAAATGTAGTCT 1080
1081 TTGGTGTAAAGAACTTCAGTTTCTGTAATTAATTTGCCCTCCGGCGAGCTGTGAACGACAC 1140
1081 TTGGTGTAAAGAACTTCAGTTTCTGTAATTAATTTGCCCTCCGGCGAGCTGTGAACGACAC 1140
1141 GTGAACGACGAGAGAAAGACAAAGAGAGTCTTCAGGATGCTTTTATTTGATTTTCTACAT 1200
1141 GTGAACGACGAGAGAAAGACAAAGAGAGTCTTCAGGATGCTTTTATTTGATTTTCTACAT 1200
1201 TTTCTTCTCTGCTGGACACCAATTTCTGTTTTTAAATACCACCAATTTTATGTTTAGGCCAA 1260
1201 TTTCTTCTCTGCTGGACACCAATTTCTGTTTTTAAATACCACCAATTTTATGTTTAGGCCAA 1260
1261 GTGACCTTTTGTAGTAAATTAAGATTTGTTTTTATGTCATGCTTATGGAACCACTATAT 1320
1261 GTGACCTTTTGTAGTAAATTAAGATTTGTTTTTATGTCATGCTTATGGAACCACTATAT 1320
1321 TFCACCTCTATTATATGATTCATCTAGACAAATTTTCAAAAGGCTTTGAAAGGTAATAA 1380
1321 TFCACCTCTATTATATGATTCATCTAGACAAATTTTCAAAAGGCTTTGAAAGGTAATAA 1380
1381 TGAAGGAGGAGTGTGTTCTTATAGTAGAGCTGATCCCTGCCTAATAATGCTGTAATAC 1440
1381 TGAAGGAGGAGTGTGTTCTTATAGTAGAGCTGATCCCTGCCTAATAATGCTGTAATAC 1440

repeat_region	/rpt_family="L1" 5317. .5493 /rpt_family="L1" 7526. .7555 /rpt_family="L1" complement(8884. .9083) /rpt_family="L1" 9114. .9199 /note="match to human EST R05512 (NID:g756132) ye92e05.s1" complement(9688. .9980) /note="match to human EST R05619 (NID:g756239) ye92e05.r1"
misc_feature	12598. .12886 /rpt_family="ALU"
repeat_region	14708. .14768 /rpt_family="L1" complement(15481. .15539) /rpt_family="L1"
misc_feature	complement(15834. .15947) /note="match to human EST AA094108 (NID:g1639701)" /rpt_family="ALU"
repeat_region	complement(15948. .16239) /rpt_family="ALU"
misc_feature	complement(18000. .18131) /note="match to human EST AAL88905 (NID:g1775949) zp79f02.r1" complement(18000. .18131) /note="match to human EST AA206526 (NID:g1801907) zq58c01.r1"
repeat_region	complement(18508. .18770) /rpt_family="ALU"
repeat_region	19013. .19068 /rpt_family="L1"
repeat_region	complement(20234. .20523) /rpt_family="ALU"
repeat_region	21100. .21124 /rpt_family="L1"
repeat_region	21172. .21342 /rpt_family="L1"
repeat_region	complement(21174. .21536) /rpt_family="L1"
repeat_region	21731. .21764 /rpt_family="L1"
repeat_region	23854. .23898 /rpt_family="L1"
repeat_region	complement(24280. .24523) /rpt_family="ALU"
repeat_region	24969. .24996 /rpt_family="L1"
repeat_region	complement(25073. .25365) /rpt_family="ALU"
repeat_region	complement(25967. .26258) /rpt_family="ALU"
repeat_region	26790. .27002 /rpt_family="L1"
repeat_region	27007. .27161 /rpt_family="ALU"
repeat_region	27171. .27273 /rpt_family="ALU"
repeat_region	27593. .27650 /rpt_family="L1"
repeat_region	27831. .28035 /rpt_family="ALU"
repeat_region	28177. .28205 /rpt_family="L1"
repeat_region	28909. .28986 /rpt_family="L1"
repeat_region	29056. .29327 /rpt_family="L1"
repeat_region	29372. .29429 /rpt_family="L1"
repeat_region	29489. .29780 /rpt_family="ALU"
repeat_region	31289. .31368 /rpt_family="THE"

[illegible]


```

Db 684 ATACCTGGGAAACACAGACACTTTATGTCTCAGTACAAATGAATACTACACATGAACGTGG 743
QY 859 GAATGTATTATACCTGTTAGTACAGATCCCAATATTTCTTTTTCACCTGTTGTAGTAAATGT 918
Db 744 GAATGTATTATACCTGTTAGTACAGATCCCAATATTTCTTTTTCACCTGTTGTAGTAAATGT 803
QY 919 TAATCATATACACCAAAATACCTTCAGCTCTTAATATTCGAATAGGCACCAAGATTTCACAA 978
Db 804 TAATCATATACACCAAAATACCTTCAGCTCTTAATATTCGAATAGGCACCAAGATTTCACAA 863
QY 979 CAGGCGAAGAAAGAAAGCAAGAAAGAAAGCAATTTCTTAACACACACATCAGG 1038
Db 864 CAGGCGAAGAAAGAAAGCAAGAAAGAAAGCAATTTCTTAACACACACATCAGG 923
QY 1039 CTACAGACATGTCCAAAGCAGTGGTGGGAGAAATGTAGTCTTTGGTGTGAAGAACTTCAG 1098
Db 924 CTACAGACATGTCCAAAGCAGTGGTGGGAGAAATGTAGTCTTTGGTGTGAAGAACTTCAG 983
QY 1099 TTTCTGTAATTAATGGCTCCGGGAGCTGTGAACACACACCGTGAACGACGAGAAAGAC 1158
Db 984 TTTCTGTAATTAATGGCTCCGGGAGCTGTGAACACACACCGTGAACGACGAGAAAGAC 1043
QY 1159 AAAGACAGCTTCAGGATCTTTATTGATATTTCTACATTTCTCTCTGCTGGACAC 1218
Db 1044 TAAGAGAGCTTCAGGATCTTTATTGATATTTCTACATTTCTCTCTGCTGGACAC 1103
QY 1219 CAATTTCTGTTTAAATPACCAACCATTTTATGTTTATGAGCCCAAGTGACCTTTTATAGTAAAT 1278
Db 1104 CAATTTCTGTTTAAATPACCAACCATTTTATGTTTATGAGCCCAAGTGACCTTTTATAGTAAAT 1163
QY 1279 TAAGATTTGTTTTTATGCTATGCTATGGAACAACTATATTTTCCACCTCTATATATG 1338
Db 1164 TAAGATTTGTTTTTATGCTATGCTATGGAACAACTATATTTTCCACCTCTATATATG 1223
QY 1339 CATTTACTACACAAAATTTTCAAAGCTTTGAAAGCTTAAATGAAAGCAAGTCTGTTT 1398
Db 1224 CATTTACTACACAAAATTTTCAAAGCTTTGAAAGCTTAAATGAAAGCAAGTCTGTTT 1283
QY 1399 CTATAGTAAAGCTGATGCCCTGCTTAATATGCTGTAAATACACAACTCTTGGATAGATC 1458
Db 1284 CTATAGTAAAGCTGATGCCCTGCTTAATATGCTGTAAATACACAACTCTTGGATAGATC 1343
QY 1459 CCAAGAAACAAAATAATTTACCTTTCAAGATAGTGAATTAAGAGAAAACGTTTAGTGC 1518
Db 1344 CTAAGAAACAAAATAATTTACCTTTCAAGATAGTGAATTAAGAGAAAACGTTTAGTGC 1403
QY 1519 CTCAGGTTGTACAGACTAGAGAAAAGCTTCAGTTTCCACCAATCCACATTCACATCAGT 1578
Db 1404 CTCAGGTTGTACAGACTAGAGAAAAGCTTCAGTTTCCACCAATCCACATTCACATCAGT 1463
QY 1579 TTTAAATTTAAATTTGTAAGAACTGATATTACTGCGCAATATAAGAAAATATTTTAAAGTA 1638
Db 1464 TTTAAATTTAAATTTGTAAGAACTGATATTACTGCGCAATATAAGAAAATATTTTAAAGTA 1523
QY 1639 TTGTTTATGTTGTAATTTTCAATGCTCAATGCTAATTTAGATAGCTCATATATTTCAAT 1698
Db 1524 TTGTTTATGTTGTAATTTTCAATGCTC-ATGTCATTTAGATAGGTCATATATTTCAAT 1582
QY 1699 TTTCTTCAATTTAAATGATTTGTTGATGCGCAGTTTGTAAAGTACTATCATGCTGATA 1758
Db 1583 TTTCTTCAATTTAAATGATTTGTTGATGCGCAGTTTGTAAAGTACTATCATGCTGATA 1642
QY 1759 TTTTGTCAATTTATGTTCCACAGAAATATTTCAATGTAAGTCAATTTTAAAGGAATAA 1818
Db 1643 TTTTGTCAATTTATGTTCCACAGAAATATTTCAATGTAAGTCAATTTTAAAGGAATAA 1702
QY 1819 ATACATAGCCTTAAACAGGTATAACTTTAAATGTAAAAAATAA 1865
Db 1703 ATACATAGCCTTAAACAGGTATAACTTTAAATGTAAAAAATAA 1749

```

RESULT 5

```

AC096185
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

```

```

AC096185      83627 bp      DNA      linear      HTG 11-JUL-2002
Rattus norvegicus clone CH230-11G1, *** SEQUENCING IN PROGRESS ***,
43 unordered pieces.
AC096185
AC096185.3   GI:21723323
HTG: HTGS-PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 83627)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R.,
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Donthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,B., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 83627)
Worley,K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 83627)
Worley,K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17943871.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

```

----- Project Information
Center project name: GBPL
Center clone name: CH230-11G1
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 31803 bases at least Q40
Consensus quality: 35840 bases at least Q30
Consensus quality: 39549 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1247: contig of 1247 bp in length
* 1248: gap of unknown length
* 1348: contig of 1062 bp in length
* 2410: gap of unknown length
* 2510: gap of unknown length
* 4041: contig of 1532 bp in length
* 4042: gap of unknown length
* 4141: gap of unknown length
* 4142: contig of 1365 bp in length
* 5507: gap of unknown length
* 5606: gap of unknown length
* 5607: contig of 1412 bp in length
* 7018: gap of unknown length
* 7019: gap of unknown length
* 7119: gap of unknown length
* 8532: contig of 1413 bp in length
* 8632: gap of unknown length
* 9748: contig of 1117 bp in length
* 9849: gap of unknown length
* 9849: contig of 1131 bp in length
* 10980: gap of unknown length
* 11079: gap of unknown length
* 12217: contig of 1138 bp in length
* 12318: gap of unknown length
* 13744: contig of 1427 bp in length
* 13745: gap of unknown length
* 13845: contig of 1908 bp in length
* 15752: gap of unknown length
* 15753: gap of unknown length
* 15853: contig of 1444 bp in length
* 17297: gap of unknown length
* 17397: contig of 1400 bp in length
* 18897: gap of unknown length
* 18897: contig of 1302 bp in length
* 20199: gap of unknown length
* 20298: gap of unknown length
* 20389: contig of 1740 bp in length
* 22039: gap of unknown length
* 22139: contig of 1762 bp in length
* 23901: gap of unknown length
* 24001: contig of 1040 bp in length
* 25041: gap of unknown length
* 25141: contig of 1317 bp in length
* 26457: gap of unknown length
* 26557: gap of unknown length
* 26558: contig of 1140 bp in length
* 27697: gap of unknown length
* 27798: contig of 1374 bp in length
* 29172: gap of unknown length
* 29271: gap of unknown length
* 29272: contig of 1050 bp in length
* 30322: gap of unknown length
* 30422: contig of 1471 bp in length
* 31892: gap of unknown length
* 31893: contig of 2002 bp in length
* 31993: gap of unknown length
* 33995: contig of 2007 bp in length
* 34095: gap of unknown length
* 36102: contig of 1196 bp in length
* 36202: gap of unknown length
* 37398: contig of 1798 bp in length
* 37498: gap of unknown length
* 39295: contig of 1798 bp in length
* 39296: gap of unknown length

39396 41152: contig of 1757 bp in length
41153: gap of unknown length
41252: contig of 2813 bp in length
44065: gap of unknown length
44165: gap of unknown length
45876: contig of 1711 bp in length
45877: gap of unknown length
45977: contig of 1876 bp in length
47852: gap of unknown length
47853: contig of 2020 bp in length
49973: gap of unknown length
50072: contig of 2729 bp in length
52801: gap of unknown length
52901: contig of 2474 bp in length
55375: gap of unknown length
55376: contig of 2569 bp in length
58044: gap of unknown length
58144: gap of unknown length
60176: contig of 2032 bp in length
60276: gap of unknown length
62792: contig of 2516 bp in length
62892: gap of unknown length
65598: contig of 2706 bp in length
65599: gap of unknown length
67840: contig of 2142 bp in length
67841: gap of unknown length
67941: contig of 2646 bp in length
70587: gap of unknown length
70687: contig of 2819 bp in length
73506: gap of unknown length
73606: contig of 2531 bp in length
76137: gap of unknown length
76237: contig of 3935 bp in length
80171: gap of unknown length
80271: contig of 3356 bp in length
80272: Location/Qualifiers
1.83627
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-11G1"
BASE COUNT 23279 a 15508 c 16123 g 24487 t 4230 others
ORIGIN
Query Match 73.8%; Score 1388.8; DB 2; Length 83627;
Best Local Similarity 86.8%; Pred. No. 7.9e-232;
Matches 1622; Conservative 0; Mismatches 232; Indels 14; Gaps 8;
QY 1 GTATTCTTCAAGGAAACACAAATTTCTTTATATCAAAACAATCAAACTTGATG 60
DB 60431 GTATTCTTCAAAATACAACA-ATTCCCTTTACCAAAACAATCGACCCCTCATG 60489
QY 61 GTTCTTAATCTACATTTTCTTATTAATAGTTTACAACTTAAAAATTAACCTAAGTACAC 120
DB 60490 TTTCTCAATCTACATTTTGCATCAGCAGTTTACAAAATACAAAATAAATACACAAAT 60549
QY 121 AATTCAAGATTTTCTTCTTACAAAGAACAGTTTATACGTCATTTAAATGCCAAATA 180
DB 60550 AA--AAAGATATCTTGTGTTTACAAAGAACATCGTGTCTTATATCATCTATAAAGTCGC 60607
QY 181 TCAATATAGTTTATCTTATTTCTACTTTCTAGGAAAAAACCACTGCTCCAAAGATCT 240
DB 60608 AAACATCAAAATATCTTATTTGATTTTCTAGGG-GAATATCCCACTGCTCCAAAGATCT 60666
QY 241 GTTTTCTCCCAATCTCGAAATCAACATGCAGTCTGAATCTAACATTAACAGTCCGAGATG 300
DB 60667 GTTTTCTCCCTGTTCTGGAATCAACATGCAGTCTGAATCAACATCAACGTCGAGATG 60726
QY 301 ACATTCATGATCAATCAACACCAATATGTACCAACCACTATCATATCGGTTAAGCTTTCAAG 360
DB 60727 ACATTCGAGCATCGATACCAATATGTACCAACCACTGTCATACCAATTAAGCTTTCAAG 60786
QY 361 TGTCTCTCACCGGATTTCTTATGTTAGAAATGTTGGAGCTTGGCAGCAACCTCACATG 420
DB 60787 TGTCTCTCACGGAATTTCTCATGTTAGAAATGTTGCTGGGGCTTGGTAGCAACCTTACCG 60846

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L25186

Center Clone name: 181_E_5

* NOTE: This record contains 91 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 691 790: contig of 690 bp in length
* 791 790: gap of 100 bp
* 1551 1551: contig of 761 bp in length
* 1552 1651: gap of 100 bp
* 1652 2395: contig of 744 bp in length
* 2396 2495: gap of 100 bp
* 2496 3216: contig of 721 bp in length
* 3217 3316: gap of 100 bp
* 3317 4046: contig of 730 bp in length
* 4047 4146: gap of 100 bp
* 4147 4877: contig of 731 bp in length
* 4878 4977: gap of 100 bp
* 4978 5716: contig of 739 bp in length
* 5717 5816: gap of 100 bp
* 5817 6556: contig of 740 bp in length
* 6557 6656: gap of 100 bp
* 6657 7387: contig of 731 bp in length
* 7388 7487: gap of 100 bp
* 7488 8245: contig of 758 bp in length
* 8246 8345: gap of 100 bp
* 8346 9066: contig of 721 bp in length
* 9067 9166: gap of 100 bp
* 9167 9880: contig of 714 bp in length
* 9881 9980: gap of 100 bp
* 9981 10728: contig of 748 bp in length
* 10729 10828: gap of 100 bp
* 10829 11539: contig of 711 bp in length
* 11540 11639: gap of 100 bp
* 11640 12371: contig of 732 bp in length
* 12372 12471: gap of 100 bp
* 12472 13190: contig of 719 bp in length
* 13191 13290: gap of 100 bp
* 13291 14034: contig of 744 bp in length
* 14035 14134: gap of 100 bp
* 14135 14890: contig of 756 bp in length
* 14891 14990: gap of 100 bp
* 14991 15717: contig of 727 bp in length
* 15718 15817: gap of 100 bp
* 15818 16569: contig of 752 bp in length
* 16570 16669: gap of 100 bp
* 16670 17396: contig of 727 bp in length
* 17397 17496: gap of 100 bp
* 17497 18249: contig of 753 bp in length
* 18250 18349: gap of 100 bp
* 18350 19051: contig of 702 bp in length
* 19052 19151: gap of 100 bp
* 19152 19903: contig of 752 bp in length
* 19904 20003: gap of 100 bp
* 20004 20736: contig of 733 bp in length
* 20737 20836: gap of 100 bp
* 20837 21567: contig of 731 bp in length
* 21568 21667: gap of 100 bp
* 21668 22389: contig of 722 bp in length
* 22390 22489: gap of 100 bp

* 22490 23215: contig of 726 bp in length
* 23216 23315: gap of 100 bp
* 23316 24044: contig of 729 bp in length
* 24045 24144: gap of 100 bp
* 24145 24885: contig of 741 bp in length
* 24886 24985: gap of 100 bp
* 24986 25745: contig of 760 bp in length
* 25746 25845: gap of 100 bp
* 25846 26702: contig of 757 bp in length
* 26703 27460: contig of 758 bp in length
* 27461 27560: gap of 100 bp
* 27561 28320: contig of 760 bp in length
* 28321 28420: gap of 100 bp
* 28421 29134: contig of 714 bp in length
* 29135 29234: gap of 100 bp
* 29235 29988: contig of 754 bp in length
* 29989 30088: gap of 100 bp
* 30089 30801: contig of 713 bp in length
* 30802 30901: gap of 100 bp
* 30902 31622: contig of 721 bp in length
* 31623 31722: gap of 100 bp
* 31723 32440: contig of 718 bp in length
* 32441 32540: gap of 100 bp
* 32541 33265: contig of 725 bp in length
* 33266 33365: gap of 100 bp
* 33366 34119: contig of 754 bp in length
* 34120 34219: gap of 100 bp
* 34220 34973: contig of 754 bp in length
* 34974 35073: gap of 100 bp
* 35074 35834: contig of 761 bp in length
* 35835 35934: gap of 100 bp
* 35935 36689: contig of 755 bp in length
* 36690 36789: gap of 100 bp
* 36790 37540: contig of 751 bp in length
* 37541 37640: gap of 100 bp
* 37641 38391: contig of 751 bp in length
* 38392 38491: gap of 100 bp
* 38492 39239: contig of 748 bp in length
* 39240 39339: gap of 100 bp
* 39340 40095: contig of 756 bp in length
* 40096 40195: gap of 100 bp
* 40196 40890: contig of 695 bp in length
* 40891 40990: gap of 100 bp
* 40991 41719: contig of 729 bp in length
* 41720 41819: gap of 100 bp
* 41820 42554: contig of 735 bp in length
* 42555 42654: gap of 100 bp
* 42655 43361: contig of 707 bp in length
* 43362 43461: gap of 100 bp
* 43462 44203: contig of 742 bp in length
* 44204 44303: gap of 100 bp
* 44304 45057: contig of 754 bp in length
* 45058 45157: gap of 100 bp
* 45158 45898: contig of 741 bp in length
* 45899 45998: gap of 100 bp
* 45999 46762: contig of 764 bp in length
* 46763 46862: gap of 100 bp
* 46863 47626: contig of 764 bp in length
* 47627 47726: gap of 100 bp
* 47727 48445: contig of 719 bp in length
* 48446 48545: gap of 100 bp
* 48546 49261: contig of 716 bp in length
* 49262 49361: gap of 100 bp
* 49362 50115: contig of 754 bp in length
* 50116 50215: gap of 100 bp
* 50216 50931: contig of 716 bp in length
* 50932 51031: gap of 100 bp
* 51032 51765: contig of 734 bp in length
* 51766 51865: gap of 100 bp
* 51866 52589: contig of 724 bp in length
* 52590 52689: gap of 100 bp
* 52690 53433: contig of 744 bp in length

```
* 53434 53533: gap of 100 bp
* 53534 54285: contig of 752 bp in length
* 54286 54385: gap of 100 bp
* 54386 55142: contig of 757 bp in length
* 55143 55242: gap of 100 bp
* 55243 55990: contig of 748 bp in length
* 55991 56090: gap of 100 bp
* 56091 56846: contig of 756 bp in length
* 56847 56946: gap of 100 bp
* 56947 57705: contig of 759 bp in length
* 57706 57805: gap of 100 bp

Query Match      28.5%; Score 535.8; DB 2; Length 76186;
Best Local Similarity 88.7%; Pred. No. 2e-80;
Matches 636; Conservative 0; Mismatches 74; Indels 7; Gaps 5;

QY 492 GTACTTGATGTAATAATTTGTGGGATGATTCCTCTAACT-ATAGTATTCCTTCGTCT 550
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70265 GAACNCGATGCATAAATTTGTGGGATGATTCCTCTAACTNATAGTATTCGTCT 70206

QY 551 -TTCACATGGAGAGTAACACACGCTCTCAAT-TGCGTTTCCATGAGGCTGTGTATCCTTT 608
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70205 CNTCACATGGAGAGTAACACGCTCTCAATTCGTCTGTTCACGAAAGCTGTGTTCCTTT 70146

QY 609 GCAAGTCTCTCAACAGCAATCAAGTTTTCGTATCACTTTGGACAGATGACATCTCT 668
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70145 GCAAGTCTTCGACAGCAATCAAGTTTTCGTAT-TCTCTGGACAGATGACATCTCT 70087

QY 669 GTAAACCTGCAACCGAATTTGACAAATGGCAGAGCTGTAATGTTAAATGATATCCATT 728
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70086 GTAAACCTGCAACAGCAATTTGACAAATGGCAGAGCTGTAATGTTAAATGATATCCATT 70027

QY 729 TGGATTTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 788
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70026 TGGATTTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 69967

QY 789 CAAAGTGGAAATACCTGGGAAACAAACACACTTTTATGTGTCAGTACAAATGAATACAC 848
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69966 CAAAGTGGAAATACATGGGAAACAAACACACTGCTGTGTCAGTACAAATGAATACAT 69907

QY 849 ACTGAACCTGGGAATGATATATCACTCTGTATGACAGATCCCAATATTTCTTTTCACTGTT 908
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69906 ACTGAGCTCGGAGTACTATCACTCTTTTGTGTCAGATCCCACTCTCTCTTTCACAGTT 69847

QY 909 GTAGTAATGTTAATCACATACACAAATACTTCAGGCTCTTATATTCGAATAGGCACA 968
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69846 ATAGTCATGTTGATCACATACACTTAAGTACTCCAGGCTCTTAACATCCGATAGGCAC 69787

QY 969 AGATTTTCACAGGCGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1028
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69786 AGATTTTCACAGGCGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 69728

QY 1029 CAACATGAGGCTACAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTGGTGTA 1088
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69727 --ACATGAGCACCAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTGGTG 69670

QY 1089 AGAATTTTCAAGTTCTGTAATAATTTGCCCTCGGCGAGCTGTGAACACGACCCGTAACGA 1148
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69669 AGAATTTTCAAGTTCTGTAATAATTTGCCCTCGGCGAGCTGTGAACACGACCCGTAACGA 69610

QY 1149 CGAAGAACACAAAGAGAGCTTTCAGATGCTTTTATGATATTTCTACATTTCTT 1205
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69609 CGAAGAACACAAAGAGAGCTTTCAGATGCTTTTATGATATTTCTACATTTCTT 69553

RESULT 7
AC116548/c
LOCUS
DEFINITION
  Dictyostellium discoideum chromosome 2 map 2886184-2983865 strain
  AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION
  AC116548
VERSION
  AC116548.1 GI:19807722
KEYWORDS
  HTG; HTGS_PHASE2.
```

```
SOURCE
ORGANISM
  Dictyostellium discoideum.
REFERENCE
AUTHORS
  1 (bases 1 to 97683)
    Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
    Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
    Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
    Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
    Noegel,A.A.
TITLE
  Sequence and Analysis of Chromosome 2 of Dictyostellium
JOURNAL
REMARK
  The Dictyostellium Genome Sequencing Consortium
REFERENCE
AUTHORS
  Baumgart,C.
TITLE
  Direct Submission
COMMENT
  Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
  Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
  CDS predictions from GeneID may contain errors. Further Information
  is available from IMB Jena, Department of Genome Analysis
  (http://genome.imb-jena.de/dictyostellium/)
  and the University Cologne, Institute for Biochemistry I
  (http://www.uni-koeln.de/dictyostellium/project.shtml)
  Funding
  Agency : Deutsche Forschungsgemeinschaft (DFG).
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.

FEATURES
      source
      1..97683
         /organism="Dictyostellium discoideum"
         /strain="AX4"
         /db_xref="taxon:44689"
         /chromosome="2"
         /map="2886184-2983865"
         /complement(103..1243)
         /note="ORF_ID:dd_01895"
         /pseudo
         /codon_start=1
         complement(join(2308..2721,2817..3046,3148..3526,3594))
         /note="ORF_ID:dd_01893"
         /pseudo
         complement(4441..4788)
         /note="ORF_ID:dd_01891"
         /codon_start=1
         /product="hypothetical protein"
         /protein_id="AAL99295.1"
         /db_xref="GI:19807723"
         /translation="MPHTSVKTKTNTNQQKYIENWEGYEHVKLIKLRPMKYKCSN
         SLKRICGGEELIHTLSAMNGNTYKIHVIDPSTPVYNLKRHNLSAYITRKNFS
         PFILVEKTKDDLFM"
         complement(join(5076..6027,6292..6302))
         /note="ORF_ID:dd_01888"
         /pseudo
         /codon_start=1
         complement(join(7498..17492,17678..17936))
         /note="ORF_ID:dd_03234"
         /codon_start=1
         /product="AT1-46.11/98"
         /protein_id="AAL99296.1"
         /db_xref="GI:19807724"
         /translation="MKDIQHILNGNSPKRVOKLASHIPSTIVOKYIMKLPQDERAP
         DISISMPLVICIFYSIAGSYSCQSNYLKFLBESFRIIAEAGDGLKFNFGEEILAV
         WYNSDKESGEQIFSLRRATRCQAQIRAMFYLISKEENGKTELFVNSNNINNNIN
         INNNNNNNNNNNSGNSFNINNNNTNEOSYSSNPITPSSSSNINIOHSDYS
         FTPOFNMTPTISSPIAKSNPSFRERVEIEQKEREKREKLNINTSSNKERFKOQNV
         QTSISPPSPSQFYSQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO
         SSFSTNSGININISNNQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO
         LTGDGFLGYNMNRNVEIASGQIFPLQLSLKSKUSKDFSKLSGVSNVDNNNS
         NNNNNNTVPISPVSNFISGVSNLLGYLNKONKONKDKDKSDNNNNNNNNNN
         NVKNSHLFSQIIVNRLVSLIKSHSQCEVFEDSDSLFSQDNDNNNSNITSIKNT
         GTKIETTEITATATTDDLELGFVKNIDYEFYIFKGMHDVSKVPFPMDKTLEELS
         PEILIQSLLMCHIPHIRVLVTDSTSEYIFLEGHKKLKKDKYLNKQNSLLNSI
```



```
Db 113022 ATATATAATATAATATAATATATTTTATATATTTATATATAGATAATATATATTTATAT 112963
Qy 1692 ATTCAATTTCTTCAATTAATGTAATTTGTCATGCCAGCTTGTAAAGACTATCAT 1751
Db 112962 ATTTATATATTTATATATAATATAATATTTTATATATTTATATATTTATATATATATTTT 112903
Qy 1752 GTGTATATTTTGTCAATATTAATGTCACACAGAAATATTCATGTAAGTCATATTTTAA 1811
Db 112902 ATATATATTTTATATATATTTTATATATAATACATATTTTATACATATTTTATATATAT 112843
Qy 1812 GGAATAAATACATAGCCCTTAAACAGTGTATAACCTTTTAAATGTAACAAAAA 1871
Db 112842 ATATTTTATATATTTTATATATATATATATTTTATATATATATATATATATATATAT 112783
Qy 1872 AAAAAA 1881
Db 112782 ATATATAATA 112773

RESULT 10
AC116967
LOCUS AC116967 93491 bp DNA linear HTG 04-APR-2002
DEFINITION Dictyostelium discoideum chromosome 2 map 5401525-5495014 strain
AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC116967
VERSION AC116967.1 GI:19920066
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 93491)
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
Noegel,A.A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL Unpublished
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 93491)
AUTHORS Baumgart,C.
JOURNAL Direct Submission
COMMENT Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
CDS predictions from GeneID may contain errors. Further Information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. 93491
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="5401525-5495014"
BASE COUNT 36436 a 9904 c 10019 g 37032 t 100 others
ORIGIN
Query Match 4.0%; Score 75; DB 2; Length 93491;
Best Local Similarity 47.9%; Pred. No. 0.0016;
Matches 245; Conservative 0; Mismatches 265; Indels 1; Gaps 1;
Qy 1372 AAAGTAAATGAAGCGAGTGTCTTATAGTAGAGCTGCCCTGCCTAATAATG 1431
Db 44473 ATACTTATATATATATATATATTTATTTGAATATAGAGATATAAAATTTTATATATATAAA 44532
Qy 1432 CTGTATACACACTCTTGGATAGATCCCAAAAGNAACAAAAATTTACCTTTGAGATA 1491
```

```
Db 44533 TTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 44592
Qy 1492 GTGAAATGAAGAGAAAACGTTTGTAGCCCTCAGGTTGTCACAGACTAGAGAAAAGTCTCAG 1551
Db 44593 AAAAAAACAACAAAATAATAAGGTGGGGAAAAATTTTTTTTCTGTTTTTAAACTCGTT 44652
Qy 1552 TTTCCACCAATCCCATTCATCAATGAGTTTAAATTTAAATTTGTAATAAACTGATATTAATCG 1611
Db 44653 TTTTAAAAAATAATAGATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 44712
Qy 1612 CCAATATAAGAAAAATATTTTAAAGTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1670
Db 44713 TTATTAGTAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 44772
Qy 1671 CTAATTAGTAGGTATATATATTTCAATTTCTTCATTAATGTTGTTGTTGTTGTTGTTGTTG 1730
Db 44773 TTTATGATCTATTTGATCTTATTTAAATTTTAAATTTTATTTTATTTTATTTTATTTTAT 44832
Qy 1731 AGTTTGTAAAGTACTATCATCTGTTATATTTTGTCAATATTTATGTCACACAGAAATATTT 1790
Db 44833 TGTAAATTTTCTTAATTTATATTTTTCACAAAGTCACAAATCTTAAACTCAAAATTTT 44892
Qy 1791 CATGTAAGTCATATTTTAAAGGAATAATAATACATAGCCTTAAACAGTGTATAACTTTAA 1850
Db 44893 CATCAGCATTAATTAAGATGGTGACTGTGGAATAATAATAATAATAATAATAATAATA 44952
Qy 1851 AATGTAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1881
Db 44953 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 44983

RESULT 11
AX344555/c
LOCUS AX344555 34980 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 6 from Patent WO0200932.
ACCESSION AX344555
VERSION AX344555.1 GI:18492441
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of known genetic parameters within the mhc
JOURNAL Patent: WO 0200932-A 6 03-JAN-2002;
FEATURES
Epigenomics AG (DE)
Location/Qualifiers
1. 34980
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo
sapiens)-Original length of seq 1: 3.673778 <223>-split as
follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001
649.980-seq 03 600.001 949.980-seq 04 900.001
1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001
1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001
2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001
3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001
3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length
of seq 2: 3.673778 <223>-split as follows.-seq 14
0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16
600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001
1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001
2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001
2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001
3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001
3.673.778"
BASE COUNT 95599 a 4385 c 78257 g 171739 t
ORIGIN
Query Match 4.0%; Score 74.4; DB 6; Length 349980;
Best Local Similarity 47.1%; Pred. No. 0.0015;
```

Matches	260;	Conservative	0;	Mismatches	291;	Indels	1;	Gaps	1;
QY	1331	ATTATATGCAFTCTACTAGACACAAAATTTCAAAAGGTCTTGAAGAAAGTAAATGAAAAAGCG	1390						
Db	214154	AATATATATATAAAATATATATAAAATATATATAAAATATATATAAAATATATATAAA	214095						
QY	1391	AGTTGTTCTTACTAGAGCTGATCCCTGCCCTAAATAAGTCTTAATACACAACTCTTG	1450						
Db	214094	ATATATATATAAAATATATATAAAATATATATAAAATATATATAAAATATATATAAA	214035						
QY	1451	GATAGATCCCAAAAGAAACAAAAAATTTACCTTTGAAGATAGTGCAAAATAGAGAAAAACG	1510						
Db	214034	TATATATATAAAATATATATAAAATATATATAAAATATATATAAAATATATATAAA	213975						
QY	1511	TTTAGTGCCTCAGGTTGTCACAGACTAGAGAAAAGTCTCAGTTTCACCAAAATCCACATTC	1570						
Db	213974	ATATATATAAAATATATATAAAATATATATAAAATATATATAAAATATATATAAA	213915						
QY	1571	AAATGAGTGTTTAAATTTAAATTTGTAAGAACTGATATTACTGCCAAATATATAAGAAAAATAT	1630						
Db	213914	TATATATAAAATATATAAAATATATATAAAATATATATAAAATATATATAAA	213855						
QY	1631	TTTAAGTATTGGTATGCT-TGTAATTTTCAAGTGTGAAATGCTTAATAGATAGTGCATAT	1689						
Db	213854	AATATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAA	213795						
QY	1690	ATATTCAATTTCTTCATCTACTTAATGTATTTGGTGGCATGGCAGTGTCTTGAAGTACTATC	1749						
Db	213794	ATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAA	213735						
QY	1750	ATGTGTATATTTTGTCAATATATTATGTCACACAGAAAAATATTCATGTAAGTCATATTTTT	1809						
Db	213734	ATATAATATATATATATATATATATAATATAATATAAAATATATAAAATATATAAA	213675						
QY	1810	AAGGAATAAATACATAGCCCTTAAACAGGTGTATTAACCTTTTAAATGTAAAAAATAAA	1869						
Db	213674	ATATATAAAATATATAAAATATATAAAATATATAAAATATATAAAATATATAAA	213615						
QY	1870	AAAAAAAAAAAAA 1881							
Db	213614	AAATATATATAA 213603							
RESULT 12	AL807813/C								
LOCUS	Human DNA sequence from clone RP13-487C10 on chromosome X, complete sequence.	106958 bp	DNA	linear	PRI 26-JUL-2002				
DEFINITION	AL807813								
ACCESSION	AL807813								
VERSION	AL807813.6	GI:22002803							
KEYWORDS	HTG.								
SOURCE	Homo sapiens.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
TITLE	1 (bases 1 to 106958)								
JOURNAL	Whitehead, S.								
	Direct Submission								
	Submitted (26-JUL-2002); Wellcome Trust Sanger Institute, Hinxton,								
	Cambridgeshire, CB10 1SA, UK. E-mail enquiries:								
	humquery@sanger.ac.uk UK clone requests: clonerequest@sanger.ac.uk								
	On Jul 29, 2002 this sequence version replaced gi:21998278.								
COMMENT	----- Genome Center								
	Center: Wellcome Trust Sanger Institute								
	Center code: SC								
	Web site: http://www.sanger.ac.uk								
	Contact: humquery@sanger.ac.uk								

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> RP13-47C10 is from the library RPC1-13.3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6.

```

FEATURES
source
LOCATION: Placement
Location/Qualifiers
1..106958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-487C10"
/clone_lib="RPC1-13.3"
BASE COUNT 36290 a 20496 c 20177 g 28995 t
ORIGIN

```

[illegible]

RESULT	13
AC007158	
LOCUS	201299 bp DNA linear PRI 17-AUG-2001
DEFINITION	Homo sapiens, clone RP11-90A1, complete sequence.
ACCESSION	AC007158
VERSION	AC007158.10 GI:11610938
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 201299)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens, clone RP11-90A1
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 201299)
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.

Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeArellano, K., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
 Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
 Melldrim, J., Mollia, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Testave, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

TITLE

JOURNAL

Submitted (24-MAR-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 201299)

AUTHORS

Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A.,
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheters, R., Melldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
 Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testave, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zalnoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (12-DEC-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Dec 9, 2000 this sequence version replaced gi:11276252.
 All repeats were identified using RepeatMasker:

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L659

Center clone name: 90_A_1

FEATURES

source

Location/Qualifiers

1..201299
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-90A1"
 /clone_lib="RP11-11 Human Male BAC"

repeat_region

1664..1684

repeat_region

/rpt_family="AT-rich"

repeat_region

3468..3598

repeat_region

/rpt_family="L2"

repeat_region

3700..3758

repeat_region

/rpt_family="(CAA)n"

repeat_region

3785..3839

repeat_region

/rpt_family="L2"

repeat_region

complement(4495..5019)

repeat_region

/rpt_family="L1MA9"

repeat_region

5029..5080

repeat_region

/rpt_family="(TA)n"

5138..5268
 /rpt_family="(TA)n"
 complement(5271..5356)
 /rpt_family="L1MA9"
 5492..5518
 /rpt_family="AT-rich"
 complement(5520..5825)
 /rpt_family="AluJo"
 complement(5903..6084)
 /rpt_family="MLT1A2"
 6110..6151
 /rpt_family="GA-rich"
 complement(6168..6322)
 /rpt_family="MLT1A2"
 complement(6329..6752)
 /rpt_family="MLT1A2-int"
 complement(6898..6959)
 /rpt_family="MLT1A2-int"
 complement(7018..7078)
 /rpt_family="MLT1A2-int"
 complement(7107..7761)
 /rpt_family="MLT1A2-int"
 complement(7764..8176)
 /rpt_family="MLT1A2"
 8650..8774
 /rpt_family="MIR"
 9716..9858
 /rpt_family="CT-rich"
 10648..11054
 /rpt_family="MLT1B"
 complement(11497..11567)
 /rpt_family="MIR"
 13140..13163
 /rpt_family="AT-rich"
 complement(14648..14845)
 /rpt_family="MIR"
 15776..16075
 /rpt_family="AluSx"
 17569..17855
 /rpt_family="AluSg"
 17856..17878
 /rpt_family="(CCAA)n"
 complement(18262..18315)
 /rpt_family="L2"
 18356..18569
 /rpt_family="MERS8A"
 19351..19399
 /rpt_family="(TG)n"
 20209..20524
 /rpt_family="AluYa5"
 21099..21240
 /rpt_family="MER99"
 21414..21910
 /rpt_family="PTRS"
 21909..21951
 /rpt_family="PTRS"
 complement(22254..22465)
 /rpt_family="MER63A"
 complement(24968..25003)
 /rpt_family="MER17"
 25617..27797
 /rpt_family="L1PA7"
 29160..29208
 /rpt_family="(TAAA)n"
 31294..31333
 /rpt_family="(T)n"
 34460..34497
 /rpt_family="A-rich"
 complement(35336..35535)
 /rpt_family="MIR"
 35562..35657
 /rpt_family="GA-rich"
 35670..35751

repeat_region	/rpt_family="GA-rich" 35998. .36084 /rpt_family="MIR" complement(36446. .36608) /rpt_family="FRAM" complement(36644. .36776) /rpt_family="MIR" complement(37640. .37803) /rpt_family="MIR" complement(39017. .39404) /rpt_family="MTIA2" 40700. .41014 /rpt_family="(TA)n" 40882	ACCESSION VERSION KEYWORDS SOURCE ORGANISM
unsure		
repeat_region	41018. .41201 /rpt_family="(TATAA)n" 41207. .41268 /rpt_family="(TA)n" complement(41650. .41786) /rpt_family="MERS58B" complement(41905. .48040) /rpt_family="LLPA4" 49825. .49872 /rpt_family="AT-rich" complement(49888. .50121) /rpt_family="MLTID" 50167. .50225 /rpt_family="(TC)n" complement(50260. .50435) /rpt_family="MLTID" complement(50717. .51018) /rpt_family="AluSx" 51019. .51159 /rpt_family="MIR" 51466. .51628 /rpt_family="MER104" 51800. .51973 /rpt_family="MIR" complement(52323. .52794)	REFERENCE AUTHORS TITLE JOURNAL
repeat_region		REFERENCE AUTHORS TITLE JOURNAL
repeat_region		REFERENCE AUTHORS TITLE JOURNAL
repeat_region		REFERENCE AUTHORS TITLE JOURNAL

RESULT 14

AC010903					
LOCUS	AC010903	91723 bp	DNA	linear	PRI 07-NOV-2001
DEFINITION	Homo sapiens BAC clone RP11-550H9 from 2,				complete sequence.

AC010903	AC010903.8	GI:13162555	
VERSION	HTG.		
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 91723)		
TITLE	Sulston,J.E. and Waterston,R.		
JOURNAL	Toward a complete human genome sequence		
MEDLINE	Genome Res. 8 (11), 1097-1108 (1996)		
PUBMED	99063792		
REFERENCE	9847074		
AUTHORS	2 (bases 1 to 91723)		
TITLE	Mulvaney,E., Haakenson,W. and Belter,E.		
JOURNAL	The sequence of Homo sapiens BAC clone RP11-550H9		
REFERENCE	Unpublished		
AUTHORS	3 (bases 1 to 91723)		
TITLE	Waterston,R.H.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (25-SEP-1999) Genome Sequencing Center, Washington		
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
TITLE	4 (bases 1 to 91723)		
JOURNAL	Waterston,R.H.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (28-FEB-2001) Genome Sequencing Center, Washington		
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
JOURNAL	5 (bases 1 to 91723)		
REFERENCE	Waterston,R.		
AUTHORS	Direct Submission		
TITLE	Submitted (09-AUG-2001) Department of Genetics, Washington		
JOURNAL	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
REFERENCE	6 (bases 1 to 91723)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-NOV-2001) Department of Genetics, Washington		
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
AUTHORS	On Feb 28, 2001 this sequence version replaced gi:11465184.		
TITLE	----- Genome Center		
JOURNAL	Center: Washington University Genome Sequencing Center		
COMMENT	Web site: http://genome.wustl.edu/gsc		
	Contact: sapiens@watson.wustl.edu		
	----- Summary Statistics		
	Center project name: H_NH0550H09		

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-270E5, 200 bp overlap.
Actual start of this clone is at base position 65201 of RP11-270E5;
actual end of this clone is at base position 91723 of RP11-550H9.

FEATURES

```

source
1. .91723
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone_lib="RPCI-11"
   /map="2"
   /clone="RP11-550H9"
1. .222
   /rpt_family="MaLR"
223..432
   /rpt_family="MaLR"
446..498
   /rpt_family="MIR"
541..1053
   /rpt_family="ERVK"
1281..1315
   /rpt_family="(ATG)n"
1339..1489
   /rpt_family="MIR"
3517..3540
   /rpt_family="AT_rich"
5843..6781
   /rpt_family="ERVL"
6945..7159
   /rpt_family="ERVL"
8558..8633
   /rpt_family="MER81"
8953..8985
   /rpt_family="(CA)n"
9444..9753
   /rpt_family="Alu"
10756..10776
   /rpt_family="AT_rich"
11621..11893
   /rpt_family="L1"
11892..12180
   /rpt_family="L1"
12181..12508
   /rpt_family="Alu"
12509..12974
   /rpt_family="L1"
13192..13285
   /rpt_family="(TCTA)n"
13832..13853
   /rpt_family="AT_rich"
14983..15164
   /rpt_family="MaLR"
15171..15457
   /rpt_family="Alu"
16219..16302
   /rpt_family="CT_rich"
16315..16980
   /rpt_family="L1"
17005..17136
   /rpt_family="MaLR"
17137..17212
   /rpt_family="(GA)n"
17213..17576
   /rpt_family="MaLR"
17585..17667
   /rpt_family="L1"
17668..17933
   /rpt_family="Alu"

repeat_region
17934..18007
   /rpt_family="L1"
repeat_region
18010..18682
   /rpt_family="L1"
repeat_region
18835..18972
   /rpt_family="L1"
repeat_region
19035..19072
   /rpt_family="(T)n"
repeat_region
19105..19700
   /rpt_family="L1"
repeat_region
20133..20416
   /rpt_family="Alu"
20511..20566
   /rpt_family="(GA)n"
21757..21786
   /rpt_family="(TG)n"
21814..22123
   /rpt_family="L2"
22510..22689
   /rpt_family="MIR"
22913..22942
   /rpt_family="AT_rich"
23865..23890
   /rpt_family="AT_rich"
25786..25830
   /rpt_family="AT_rich"
25838..26143
   /rpt_family="Alu"
27275..27411
   /rpt_family="L1"
27862..28053
   /rpt_family="L1"
28055..28747
   /rpt_family="L1"
28749..28921
   /rpt_family="AT_rich"
28922..29049
   /rpt_family="(TTATA)n"
29050..29102
   /rpt_family="AT_rich"
29103..29202
   /rpt_family="(TTATA)n"
29207..29316
   /rpt_family="AT_rich"
29324..29668
   /rpt_family="L1"
29696..29845
   /rpt_family="L1"
29842..30078
   /rpt_family="L1"
30098..30125
   /rpt_family="AT_rich"

Query Match
Best Local Similarity 48.18; Pred. No. 0.0051;
Matches 296; Conservative 0; Mismatches 315; Indels 5; Gaps 3;

Qy 1264 ACCTTTAGTAAATTAAGATTGCTTTTGTAGTCATCGCTTATGGAACAACATATATTC 1323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28757 ACATATATATAAATATACATATATGATATATACATATATATATATATATATATATATAT 28816

Qy 1324 ACCCTCTATTATATGCTACTACAGACAAAATTTCAAAGGCTCTTGAAGTAAATGA 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28817 ATTATATATGATATATATATGATATATATATATATATATATATATATATATATATAT 28876

Qy 1384 AAAAGCGAGTTGTTCTCTAGTAGAGAGCTGATCCCTGCCCTAAATATCTGTAATACACA 1443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28877 TATGATATATATATATGATATA-TAATACATATATATATATATATATATATATATATAT 28935

Qy 1444 ACTCTTGATAGATCCCAAGAAACAAAAAATTTACCTTTGAAGATAGTGAATAAGAG 1503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28936 TATAATAATATATATATATATATATATATATATATATATATATATATATATATATAT 28995

```


QY	1423	CTAATAATGCTGTAATACACAACTCTTGATAGATGCCAAAAGAAACAAAAAATTACCT	1482
Db	9184	TTTGAATTTGAGCTCATACGAGCTAAATCTGATATAGTAATACTATGGATCTAATAAANAAG	9125
QY	1483	TTGAAGATAGTGAATTAAGAGAAAAACGTTTAGTGCCTCAGTTGTCACAGACTAGAGAA	1542
Db	9124	TTTAGAGTGGCTCTTTATAACCTTTTTTTTGGTGCGAATAAATAAAAAAATGAAAAAA	9065
QY	1543	AAGTCTCAGTTTACCACCAATCCACATTCMAATGAGTTTAAATTTGAAAAACTG	1602
Db	9064	AAATTTTTTTTTTAAAAATAAAAAAGAAAGAAAAATAAAAAAATTTAAAAAANA	9005
QY	1603	ATATTACTGCCAAATATAAGAAAAATATTTAAGTATTTGGTTATGTCGTAAATTTTCAAT	1662
Db	9004	AAATAAAAAANAATAAATAAAAAANAANAANAANAANAANAANAANAATAAATAAAT	8945
QY	1663	GTGAATAGCTAATTAGATAGTGCATATATATCAATTTCTCAATTACTTAATGATTGT	1723
Db	8944	CAAAATTTGGACGCTTGGGGAACCTATAAACTAAAAAGTTTTTTTTTTTTTTTTTCT	8885
QY	1723	TGCATGGCAGTTTGTTAAAGTACATCATGTTGTAATATTT--TGTCAAATATTATGTCAC	1780
Db	8884	TATATTTTTCATGATTTAAATATATAAATAATATTTTTTTTAAATATATTATATTATTAAT	8825
QY	1781	AGAAAAATTTCATCTAAGTCAATATTTTTTAAAGGAATAAATACATAGCCTTAAAAACAGT	1840
Db	8824	TATTTTTAAATATTATTTTAAATTTTTTATTTATTTTATTTTATTTTGGTTTTCCCAAAATAA	8765
QY	1841	ATAACTTTAAATGTAAAAAANAANAANAANAANAANAANAANAANAANAANA	1881
Db	8764	ATAAATTTAAAAAANAATACAAATAAATAAANAANAANAANAANAANAACAATAAA	8724

Search completed: December 10, 2002, 09:25:26
Job time : 6021 secs

[illegible]

	Query Match	3.8%;	Score 71.8;	DB 2;	Length 25117;
	Best Local Similarity	45.6%;	Pred. No. 0.0074;		
	Matches 292;	Conservative 0;	Mismatches 347;	Indels 2;	Gaps 1;
QY	1243	TTTATGCTTTAGGCCCAAGTGACCTTTAGTAAATTAAGATTGCTGTTTTAGTCATCGG	1302		
Db	9364	TCFATATTATATATATGTAATTGGTTTTAATATCTCTATTATGGTTTTTTTTTTTT	9305		
QY	1303	CTTATGGAACAACATATATTTCCACCCTCTATTATATGCATTTCACTACAGACAAAATTCAA	1362		
Db	9304	TTTTCTTTCATATTTTAAACTTTATTTCTAAATATATCGGAGAAAAAATAAA	9245		
QY	1363	AGTCTTGAAGATAAATGAANAACGGAGTGTGTTCTATAGTAGAAGCTGATCCCCGTC	1422		
Db	9244	TAATAATAATAAATAATAGATATAGGATTTATTTATTTTACTTACACAGCTAGTTAAAGT	9185		

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2002, 07:33:36 ; Search time 398 Seconds

(without alignments)
10643.238 Million cell updates/sec

Title: US-09-845-721-1

Perfect score: 1881

Sequence: 1 gttatttcttcaaaaggaaa.....aaaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB-seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1881	100.0	1881	24	Human GPR22 encodi
2	1879.4	99.9	1881	24	Cysteine protease-
3	1625	86.4	2238	20	Human G-protein co
4	1302	69.2	1302	21	Human G protein-co
5	1297.2	69.0	1302	21	DNA encoding human
6	170.2	9.0	200	24	Cysteine protease-
7	157	8.3	200	24	Cysteine protease-
8	67.6	3.6	5487	24	Human immune syste
9	65.8	3.5	7857	24	Human angiogenesis

C 10	63.8	3.4	8842	24	ABL333448	Human immune syste
C 11	63.8	3.4	8842	24	AA563334	Chemically pretrea
C 12	63.4	3.4	61020	22	AA546788	Tumour suppressor
C 13	63	3.3	11555	24	ABL32616	Human immune syste
C 14	63	3.3	11555	24	AAD28380	Human chemically t
C 15	62.8	3.3	17280	22	AA546772	Tumour suppressor
C 16	62.4	3.3	50000	22	ABL55643	Amepv genome fragm
C 17	61.8	3.3	5255	24	ABX28419	DNA transcription
C 18	61.6	3.3	18683	24	ABL54334	Chemically treated
C 19	61.6	3.3	18683	24	ABL32313	Human immune syste
C 20	61.4	3.3	18988	22	AA546342	Tumour suppressor
C 21	61.4	3.3	18988	24	ABL70204	Chemically treated
C 22	61.4	3.3	18988	24	ABL32701	Human immune syste
C 23	61.4	3.3	18988	24	ABL34509	Human metastasis a
C 24	61.2	3.3	2104	13	AAQ25273	Sequence encoding
C 25	61.2	3.3	6050	24	ABL34011	Human immune syste
C 26	61	3.2	7319	24	ABL34045	Human immune syste
C 27	60.8	3.2	5487	24	ABL33599	Human immune syste
C 28	60.8	3.2	6052	24	AA561094	Human gene regulat
C 29	60.6	3.2	5276	24	ABL32150	Human immune syste
C 30	60.4	3.2	545	22	AAH70126	Human cervical can
C 31	60.4	3.2	5447	22	AA546758	Tumour suppressor
C 32	60.2	3.2	5852	12	AAQ11710	Dictyostellium plas
C 33	60.2	3.2	10048	24	ABL70314	Chemically treated
C 34	60.2	3.2	10048	24	AA561252	Human gene regulat
C 35	60	3.2	60	24	ABN45479	Human spliced tran
C 36	60	3.2	621	22	AAH71551	Human cervical can
C 37	60	3.2	883	22	ABL15210	Human breast cance
C 38	59.8	3.2	3873	24	ABO67089	Human angiogenesis
C 39	59.8	3.2	5675	24	ABL33276	Human immune syste
C 40	59.6	3.2	6035	24	ABL33325	Human immune syste
C 41	59.6	3.2	11790	24	ABL32543	Human immune syste
C 42	59.4	3.2	10682	24	ABN80085	Human chemically m
C 43	59.2	3.1	6242	24	ABL34148	Human immune syste
C 44	59	3.1	5487	24	ABL33599	Human immune syste
C 45	59	3.1	6292	22	AA546735	Tumour suppressor

ALIGNMENTS

RESULT 1

ABA05328

ID ABA05328 standard; cDNA; 1881 BP.

XX ABA05328;

AC

27-FEB-2002 (first entry)

DT

XX

DE Human GPR22 encoding cDNA.

XX

KW Human; GPR22; anorectic; appetite control agent; GPR; gene;

KW G protein-coupled receptor; orphan receptor; antisense gene therapy;

KW mouse; GPR56; SS.

XX

OS Homo sapiens.

XX

FH Key

FT CDS

FT 237..1538

FT /*tag=

FT /product= "GPR22"

XX

PN WO200183550-A2.

XX

PD 08-NOV-2001.

 XX | 30-APR-2001; 2001WO-GB01874. || XX | 03-MAY-2000; 2000US-201418P. |
| XX | (ASTR) ASTRAZENECA AB. |
| XX | (ASTR) ASTRAZENECA UK LTD. |

Db 144 AAATCAACATGACGCTGTAATCTAACATTAACAGTGCAGATGACATTAATGATGACATCAACA 203
 Qy 319 CCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAGTGTCTCAACCGGATTTTC 378
 Db 204 CCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAGTGTCTCTCACCGGATTTTC 263
 Qy 379 TTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTCACTGTATTTGGTACTTTTACTGCA 438
 Db 264 TTATGTTAGAAATTTTGGGACTTGGCAGCAACCTCACTGTATTTGGTACTTTTACTGCA 323
 Qy 439 TGAATCCAACTTTAATCAACTCTGTCAGTAAACATTTTACAATGAATCTTCATGACTTTG 498
 Db 324 TGAATCCAACTTTAATCAACTCTGTCAGTAAACATTTTACAATGAATCTTCATGACTTTG 383
 Qy 499 ATGTAATTAATTTGTTGGGATGATTCCTCTAAGTATGATTAATCTTCGTTCACTGG 558
 Db 384 ATGTAATTAATTTGTTGGGATGATTCCTCTAAGTATGATTAATCTTCGTTCACTGG 443
 Qy 559 AGAGTAACACTGCTCTCATTTGCTGTTTCCATGAGGCTTGTGTATCTTTTGCAGAGTGTCT 618
 Db 444 AGAGTAACACTGCTCTCATTTGCTGTTTCCATGAGGCTTGTGTATCTTTTGCAGAGTGTCT 503
 Qy 619 CAACAGCAATCAAGCTTTTGGTATCACTTTGGCAGATATGACATCTCTGTAAAGCTG 678
 Db 504 CAACAGCAATCAAGCTTTTGGTATCACTTTGGCAGATATGACATCTCTGTAAAGCTG 563
 Qy 679 CAACAGCAATCAAGCTTTTGGTATCACTTTGGCAGATATGACATCTCTGTAAAGCTG 738
 Db 564 CAACAGCAATCAAGCTTTTGGTATCACTTTGGCAGATATGACATCTCTGTAAAGCTG 623
 Qy 739 CTTTCT 798
 Db 624 CTTTCT 683
 Qy 799 ATACCTGGGAAACCAAGACACTTTTATGTGTCTAGTACAAATGAATACACTACACTGAAGCTG 858
 Db 684 ATACCTGGGAAACCAAGACACTTTTATGTGTCTAGTACAAATGAATACACTACACTGAAGCTG 743
 Qy 859 GAATGTATATACACTGTTAGTACAGATCCCAATATTTCTTTTCACTGTTGTAGTATGT 918
 Db 744 GAATGTATATACACTGTTAGTACAGATCCCAATATTTCTTTTCACTGTTGTAGTATGT 803
 Qy 919 TAATCACATACACCAAAATCTCTCAGGCTCTTAATATTCGAATAGGCACAGATTTTCAA 978
 Db 804 TAATCACATACACCAAAATCTCTCAGGCTCTTAATATTCGAATAGGCACAGATTTTCAA 863
 Qy 979 CAGGCGAGAAGAAAGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1038
 Db 864 CAGGCGAGAAGAAAGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 923
 Qy 1039 CTACAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTGGTGTAGAACTTCAG 1098
 Db 924 CTACAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTGGTGTAGAACTTCAG 983
 Qy 1099 TTTCTGTAAATTAATTTGCTCGGAGCTGTGAACGACACCGTGAACGAGGAGAAAGAC 1158
 Db 984 TTTCTGTAAATTAATTTGCTCGGAGCTGTGAACGACACCGTGAACGAGGAGAAAGAC 1043
 Qy 1159 AAAAGAGAGTCTCAGGATGCTTTTATTTGATTTATTTCTACATTTCTCTCTCTCTCTCTCT 1218
 Db 1044 TAAAGAGAGTCTCAGGATGCTTTTATTTGATTTATTTCTACATTTCTCTCTCTCTCTCTCT 1103
 Qy 1219 CAATTTCTGTTTAAATACCACTTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 1278
 Db 1104 CAATTTCTGTTTAAATACCACTTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 1163
 Qy 1279 TAAATTTGTTTAAATACCACTTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 1338
 Db 1164 TAAATTTGTTTAAATACCACTTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT 1223
 Qy 1339 CATTCACTAGACAAATTTTCAAAAGCTTTGAAAGCTTTGAAAGCTTTGAAAGCTTTGAAAGCT 1398
 Db 1224 CATTCACTAGACAAATTTTCAAAAGCTTTGAAAGCTTTGAAAGCTTTGAAAGCTTTGAAAGCT 1283

Qy 1399 CTATAGTAGAGCTGATCCCTGCTCTAATATGCTGTAATACACAACTCTTGGATGATC 1458
 Db 1284 CTATAGTAGAGCTGATCCCTGCTCTAATATGCTGTAATACACAACTCTTGGATGATC 1343
 Qy 1459 CCAAAAGAAACAAAATAATTAACCTTTGAAGTACTGAAATAGAGAAAAACGTTTAGTGC 1518
 Db 1344 CTAAAGAAACAAAATAATTAACCTTTGAAGTACTGAAATAGAGAAAAATGTTTAGTGC 1403
 Qy 1519 CTCAGTTGTCACAGACTAGAGAAAAGTCTCAGTTTCCAAATCCCAATCCCAATGAGT 1578
 Db 1404 CTCAGTTGTCACAGACTAGAGAAAAGTCTCAGTTTCCAAATCCCAATCCCAATGAGT 1463
 Qy 1579 TTTAAATTTAAATTTGTAARAACTGATATTAAGTAAATTAAGAAAAATATTTAAAGTA 1638
 Db 1464 TTTAAATTTAAATTTGTAARAACTGATATTAAGTAAATTAAGAAAAATATTTAAAGTA 1523
 Qy 1639 TTGCTTATGTTGTAATTTTCAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1698
 Db 1524 TTGCTTATGTTGTAATTTTCAATGTAATGTAATGTAATGTAATGTAATGTAAT 1582
 Qy 1699 TTTCTCAATTAATGTAATTTGTCAGTGGCAGTTTGTAAAGTACTATCATGTTTAAAGTA 1758
 Db 1583 TTTCTCAATTAATGTAATTTGTCAGTGGCAGTTTGTAAAGTACTATCATGTTTAAAGTA 1642
 Qy 1759 TTTTGTCAATTAATGTAATTTGTCAGTGGCAGTTTGTAAAGTACTATCATGTTTAAAGTA 1818
 Db 1643 TTTTGTCAATTAATGTAATTTGTCAGTGGCAGTTTGTAAAGTACTATCATGTTTAAAGTA 1702
 Qy 1819 ATACATAGCCTTAAACAGCTGTATAAATTTAAATGTAATGTAATGTAATGTAATGTAAT 1865
 Db 1703 ATACATAGCCTTAAACAGCTGTATAAATTTAAATGTAATGTAATGTAATGTAATGTAAT 1749

RESULT 4
 AAA30613
 ID AAA30613 standard; cdna; 1302 BP.

XX AAA30613;

XX 21-AUG-2000 (first entry)

XX Human G protein-coupled receptor GPR22 cdna.

XX G protein-coupled receptor; GPCR; constitutively active;
 KW intracellular loop 3; transmembrane domain 6; drug screening;
 KW agonist; antagonist; ss.

XX Homo sapiens.

XX WO200022129-A1.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-US23938.

XX 13-OCT-1998; 98US-0170496.

XX (AREN-) ARENA PHARM INC.

XX Behan DP, Chalmers DT, Liaw CW;

XX WPI; 2000-329165/28.

XX P-PSDB; AAY90622.

XX Non-endogenous constitutively activated human G protein-coupled
 PT receptors, useful for identifying agonists for use as pharmaceutical
 PT agents

XX Example 1; Page 136-137; 341pp; English.

XX The invention relates to constitutively active, non-endogenous versions
 CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-

CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC and AAA30775-A30779). The mutant proteins of the invention contain a
CC mutation in a portion of the protein comprising intracellular loop 3
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC is substituted for an endogenous residue in IC3 at a position 16 amino
CC acids N-terminal of an endogenous proline in TM6 to form a sequence
CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC or Ala, and is preferably Lys. When the endogenous residue at this
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC The 15 amino acid stretch between the substituted amino acid and the Pro
CC may be endogenous, non-endogenous, or a mixture of endogenous and
CC non-endogenous residues. The constitutively active GPCRs are useful for
CC identifying antagonists, agonists and partial agonists for use as
CC pharmaceutical agents. The mutant proteins are also useful in research
CC settings for elucidating the roles of the receptors in normal and
CC diseased conditions. Antagonists for a particular GPCR are useful for
CC treating diseases and disorders associated with that receptor. Because
CC the novel mutant GPCRs are constitutively active, they can be used
CC directly for screening of compounds without the need for endogenous
CC ligands. The present sequence represents cDNA encoding a human wild-type
CC GPCR used in an exemplification of the invention. This was cloned and
CC subjected to site-directed mutagenesis (SDM) to generate DNA encoding
CC the corresponding mutant of the invention.

XX Sequence 1302 BP; 411 A; 245 C; 226 G; 420 T; 0 other;

Query Match 69.2%; Score 1302; DB 21; Length 1302;
Best Local Similarity 100.0%; Pred. No. 6.9e-215;
Matches 1302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 ATGTGTTTTTCCCATTCGGAATACACATGCGAGTCGATCTGAATCTAAACATTCAGTGGCA 296
DB 1 ATGTGTTTTTCTCCCATTCGGAATCAACATGCGAGTCGATCTGAATCTAAACATTCAGTGGCA 60
QY 297 GATGACATTCATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTT 356
DB 61 GATGACATTCATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTT 120
QY 357 CAAGTGTCTCTCACCAGATTTCTTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTC 416
DB 121 CAAGTGTCTCTCACCAGATTTCTTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTC 180
QY 417 ACTGATATGGTACTTTACTGCAATGAATCCAACTTAATCAACTCTGTCAGTAACATTAAT 476
DB 181 ACTGATATGGTACTTTACTGCAATGAATCCAACTTAATCAACTCTGTCAGTAACATTAAT 240
QY 477 ACAATGAATCTTCATGCTACTTGATGTAAATAATTTGTTGGGATGATTCCTCTAACTATA 536
DB 241 ACAATGAATCTTCATGCTACTTGATGTAAATAATTTGTTGGGATGATTCCTCTAACTATA 300
QY 537 GTTATCCTTCTGCTTTCACTGGAGAGTAACACTGCTCTCATTTGCTTTGCCATGAGGCT 596
DB 301 GTTATCCTTCTGCTTTCACTGGAGAGTAACACTGCTCTCATTTGCTTTGCCATGAGGCT 360
QY 597 TGTGATCTTTTGCAAGTGTCTACAGCAATCAAGTTTTTGGCTATCACTTTGGACAGA 656
DB 361 TGTGATCTTTTGCAAGTGTCTCAACAGCAATCAAGTTTTTGGCTATCACTTTGGACAGA 420
QY 657 TATGACATCTCTGTAACACCTGCAACCCGAATTCGACAAATGGCAGAGCTCTAATGTTA 716
DB 421 TATGACATCTCTGTAACACCTGCAACCCGAATTCGACAAATGGCAGAGCTCTAATGTTA 480
QY 717 ATGATATCCATTTGGATTTTTTCTTTTTTCTCTCTTCCTGATTCCTTTTATTCAGGTAAT 776
DB 481 ATGATATCCATTTGGATTTTTTCTTTTTTCTCTCTTCCTGATTCCTTTTATTCAGGTAAT 540
QY 777 TTTTTCAGTCTTCAAGTGGAAATACCTGGGAAACAAGACACTTTTATGTCAGTACA 836
DB 541 TTTTTCAGTCTTCAAGTGGAAATACCTGGGAAACAAGACACTTTTATGTCAGTACA 600
QY 837 AATGAATACTACACTGAACCTGGAATGATATTCACCTCTTGTAGTACAGATCCCAATATTC 896
DB 601 AATGAATACTACACTGAACCTGGAATGATATTCACCTCTTGTAGTACAGATCCCAATATTC 660

QY 897 TTTTTCACCTGTTGTAGTAAGTTAATACATACACCAAAATACTTCAAGCTCTTTAATATT 956
DB 661 TTTTTCACCTGTTGTAGTAAGTTAATACATACACCAAAATACTTCAAGCTCTTTAATATT 720
QY 957 CCAATAGGCACAAAGATTTTTCACAGGGGCAGAGAGAAGCAAGCAAGCAAGCAAT 1016
DB 721 CCAATAGGCACAAAGATTTTTCACAGGGGCAGAGAGAAGCAAGCAAGCAAGCAAT 780
QY 1017 TCTCTAAACCACACAAACATCAGGCTACAGACATGTCCAAAGCAGTGGTGGGAGAAATGTA 1076
DB 781 TCTCTAAACCACACAAACATCAGGCTACAGACATGTCCAAAGCAGTGGTGGGAGAAATGTA 840
QY 1077 GTCTTTGGTGTAAAGAACTTTCAGTTCTCTGTAAATTTGCCCTCCGGGAGCTGTGAAACGA 1136
DB 841 GTCTTTGGTGTAAAGAACTTTCAGTTCTCTGTAAATTTGCCCTCCGGGAGCTGTGAAACGA 900
QY 1137 CACCGTGAACGACGAGAGAAAGACAAAGACAGCTTTCAGGATGCTTTTATGATTTTCT 1196
DB 901 CACCGTGAACGACGAGAGAAAGACAAAGACAGCTTTCAGGATGCTTTTATGATTTTCT 960
QY 1197 ACATTTCTTCTCTGCTGGACACCAATTTCTGTTTAAATACCACCAATTTTATGTTTAGGC 1256
DB 961 ACATTTCTTCTCTGCTGGACACCAATTTCTGTTTAAATACCACCAATTTTATGTTTAGGC 1020
QY 1257 CCAAGTGACCTTTTAGTAAATTAAGATTTGTTTTTTTAGTCATGCTTTATGGAACAAT 1316
DB 1021 CCAAGTGACCTTTTAGTAAATTAAGATTTGTTTTTTTAGTCATGCTTTATGGAACAAT 1080
QY 1317 ATATTTCACCCCTTATATATGATTCATCTAGACAAAATTTCAAAAGCTTTGAAAAGT 1376
DB 1081 ATATTTCACCCCTTATATATGATTCATCTAGACAAAATTTCAAAAGCTTTGAAAAGT 1140
QY 1377 AAAATGAAAAGCGAGCTTCTTCTATAGTAGAAGCTGATCCCTGCCTAATAATGCTGTA 1436
DB 1141 AAAATGAAAAGCGAGCTTCTTCTATAGTAGAAGCTGATCCCTGCCTAATAATGCTGTA 1200
QY 1437 ATACACAACCTTTGGATAGATCCCAAGAGAAACAAAATAATACCTTTGAAGATAGTGAA 1496
DB 1201 ATACACAACCTTTGGATAGATCCCAAGAGAAACAAAATAATACCTTTGAAGATAGTGAA 1260
QY 1497 ATAAAGAGAAAAGCTTTAGTGCTCAGGTTGTCACAGACTAG 1538
DB 1261 ATAAAGAGAAAAGCTTTAGTGCTCAGGTTGTCACAGACTAG 1302

RESULT 5

AAA30722

ID AAA30722 standard; DNA; 1302 BP.

XX AAA30722;

XX 21-AUG-2000 (first entry)

XX DNA encoding human mutant G protein-coupled receptor GPR22 (F312K).

DE G protein-coupled receptor; GPCR; constitutively active;

KW intracellular loop 3; transmembrane domain 6; drug screening;

KW agonist; antagonist; mutant; ss.

XX Homo sapiens.

OS Synthetic.

XX WO200022129-A1.

XX 20-APR-2000.

XX 12-OCT-1999; 99WO-US23938.

XX 13-OCT-1998; 98US-0170496.

XX (AREN-) ARENA PHARM INC.

XX


```

Db      3  TTTGATTTTCTAGGCAAAACACCACTGCTCCAAAGAAATGTGTTTCTCCTGTTCTGG 62
QY      259 AAATCAACATGCGAGTCTGAATCTAAACATTAAGTGGAGATGACATTTGATGACATCAACA 318
Db      63  AAATCAACATGCGAGTCTGAATCTAAACATGACGGCGGAGATGACATTTGATGACATCGACA 122
QY      319 CCAATATGTACCAACCACTATCATATCCGTTAAGCTTTTCAAGTGTCTCTCACCGGATTTTC 378
Db      123 CCAATATGTACCAACCACTGTATCATACCCACTAAGCTTTTCAAGTGTCTCTCACCGGATTTTC 182
QY      379 TTAGTTTAGAAATGTG 395
Db      183 TCATGTTAGAGATCGTG 199

RESULT 8
ABL33598/c
ID  ABL33598 standard; DNA; 5487 BP.
XX
AC  ABL33598;
XX
DT  26-MAR-2002 (first entry)
XX
DE  Human immune system associated gene SEQ ID NO: 1571.
XX
KW  Human; Immune system disease; cytosine methylation; antiasthmatic;
KW  antiarteriosclerotic; anti-anaemic; cytosolic; nontropic;
KW  neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW  antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW  antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW  acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW  neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW  gene; ds.
XX
OS  Homo sapiens.
XX
PN  WO200200928-A2.
XX
PD  03-JAN-2002.
XX
PF  02-JUL-2001; 2001WO-EP07537.
XX
PR  30-JUN-2000; 2000DE-1032529.
XX
PR  01-SEP-2000; 2000DE-1043826.
XX
PA  (EPIG-) EPIGENOMICS AG.
XX
PI  Olek A, Piepenbrock C, Berlin K;
XX
DR  WPI; 2002-130909/17.
XX
PT  Nucleic acid comprising fragment of chemically modified gene, useful
PT  for diagnosis and treatment of diseases associated with abnormal
PT  cytosine methylation -
XX
PS  Claim 1; SEQ ID NO 1571; 32pp + Sequence Listing; German.
XX
CC  The present invention provides a number of human immune system associated
CC  genes which are modified by the methylation of cytosines. The sequences
CC  can be used in the diagnosis and treatment of immune system disorders,
CC  including eye diseases such as retinopathy, neovascular glaucoma and
CC  macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC  leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC  rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC  diseases. The present sequence is a gene of the invention.
XX
SQ  Sequence 5487 BP; 1532 A; 133 C; 1107 G; 2715 T; 0 other;
Query Match 3.6%; Score 67.6; DB 24; Length 5487;
Best Local Similarity 47.1%; Pred. No. 0.0079;
Matches 208; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

```

```

QY      1435 TAATACACAACTCTTGGATAGATCCCAAAAGAAACAAAATAATTACCTTTCAAGATAGTG 1494
Db      780 TAAATACTACTATTAACATTAAACATAACTTTCCAAAATAATTTCTATACACAATATAT 721
QY      1495 AAATAAGAGAGAAAACGTTTGTAGTGCCTCAGGTTGTGCAGAGACTAGAGAAAAGTCTCAGTTT 1554
Db      720 ACATCTCTAAAACAATATCCTAATCCATTTATATCTTCTAAAACAATAACCTAAACAA 661
QY      1555 CACCAATCCACATTCAAATGAGTTTAAATTTAAATTTGTAACAACTGATATTTACTGCCA 1614
Db      660 AATAAATTTATAAATCTATATTTATAAATTTATAAATAATTTATAAATCTATATTTATAAAT 601
QY      1615 AATATAAGAAAATATTTTAAAGTATTTGTTTAAATTTTCAATGTCAATGCTAA 1674
Db      600 TATTTATAAATTTATAAATTTATAAATTTATAAATTTATAAATTTATAAATTTATAAAT 541
QY      1675 TTAGATAGTCATATATATTTCAATTTCTTCAATTTCTTCAATTTCTTCAATTTCTTCAAT 1734
Db      540 ATATTTATATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 481
QY      1735 TGTAAAGTACTATCATGTGTATATTTTGTCAATTTATTTGTCACACAGAAAATTTCAATG 1794
Db      480 TTTAAATTTAAATTTTATTTAAATTTAAATTTAAATTTATATATTTATATATTTATATATA 421
QY      1795 TAAGTCATATTTTAAAGGATAAATACATAGCCCTTAAACAGTGTATACCTTTTAAATG 1854
Db      420 TATTTAAATTTATANAATTTTANAATTTTANAATTTTANAATTTTANAATTTTANAATTT 361
QY      1855 TAAAAAATAAAAAAATAAAAAA 1876
Db      360 TATAATTTATAAATAAACAATA 339

RESULT 9
ABO67075/c
ID  ABO67075 standard; DNA; 7857 BP.
XX
AC  ABO67075;
XX
DT  28-AUG-2002 (first entry)
XX
DE  Human angiogenesis associated polynucleotide SEQ ID NO 105.
XX
KW  Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW  inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
KW  macular degeneration; inflammatory bowel disease; Crohn's disease;
KW  antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW  antiarteriosclerotic; ds.
XX
OS  Homo sapiens.
XX
PN  WO200246454-A2.
XX
PD  13-JUN-2002.
XX
PF  06-DEC-2001; 2001WO-EPI4320.
XX
PR  06-DEC-2000; 2000DE-1061338.
XX
PA  (EPIG-) EPIGENOMICS AG.
XX
PI  Schacht O;
XX
DR  WPI; 2002-500450/53.
XX
PT  New nucleic acid fragments from chemically treated
PT  angiogenesis-associated genes, useful for determining methylation
PT  status, e.g. in diagnosis or treatment of cancer -
XX
PS  Claim 1; SEQ ID NO 105; 41pp + Sequence Listing; German.
XX
CC  The invention relates to a nucleic acid (I) comprising a segment of 18
CC  bases of chemically pretreated DNA of angiogenesis-associated genes (II)

```

CC having sequences (AB066971-AB067178) or their complements. (1), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 7857 BP; 2110 A; 218 C; 2100 G; 3429 T; 0 other;

Query Match 3.5%; Score 65.8; DB 24; Length 7857;
Best Local Similarity 45.9%; Pred. No. 0.017;
Matches 265; Conservative 0; Mismatches 307; Indels 5; Gaps 1;

QY 1305 TATGGACAACTATATTTCCCTCTATATATGATTCACCTAGACAAAAATTTCAAAG 1364
DB 4857 TATTAATAAATAATATAACCTCCATTTTAAACCAACATATTCCAAACATTTACCT 4798
QY 1365 GTCTGAAAGTAAATGAAAAAGGAGTGTCTCTATAGTAGAGCTGATCCCTGCT 1424
DB 4797 ACATTAACACTTATATATAAACCCCTCTACTATAAATACAAAAATTAACCGAAC 4738
QY 1425 AATAATGCTGTAATACACAACTCTTTGGATAGATCCCAAGAAAC-----AAAAAATTA 1479
DB 4737 GTAATAACGTATATATATTTACCTTAAAAAACTAAACAAAAAATCGCTTAACCCGAAA 4678
QY 1480 CTTTGAAGATAGTGAATAGAGAAAAAGCTTTAGTGCTCAGGTGTGCACAGACTAGA 1539
DB 4677 ACAGAAATTTAAATTAACAATAAACCCGAAATCGTACCCTACACTCCAAACCAACAAA 4618
QY 1540 GAAAGTCTCAGTTTCACCAAAATCCACATTCAAATGAGTTTAAATTTAAATTTGTAATA 1599
DB 4617 CAAACTCTCTCACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4558
QY 1600 CTGATATTACTGCCAAATATAAGAAAAATATTTTAAAGTATGTTGTTGTAATAATTTTC 1659
DB 4557 TAAATATCATCTATTACGACAATACTATATATCTTAAATAAATAATCTAATTTTATTTAT 4498
QY 1660 AATGCAATGCTAATAGTAGTAGTCATATATATTTCAATTTCTTCAATTTAATGTAAT 1719
DB 4497 TATTTATTTTCAATAACTTACTAAAAATTTTAAATAAATAATATTAATATATCTACAT 4438
QY 1720 TCTTGATGCGAGTTGTTTAAAGTACTATCATGTGTATATTTTGTCAATATATTATGTC 1779
DB 4437 TTTTATATTATTTTAAATAATTTTATTTTATTTTCTTTCTTTTATTTTCTAT 4378
QY 1780 CAGAAAAATTCATGTAAGTCATATTTTAAAGGAATAAATACATAGCCTTAAACAGTG 1839
DB 4377 TTTAATATCTATATTTTATATATATTTTAAAAAATAAATAAATAAATAAATAAATAA 4318
QY 1840 TATACTTTAAATGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1876
DB 4317 TATAACACAATACTTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4281

RESULT 10
ABL33448/C
ID ABL33448 standard; DNA; 8842 BP.
XX ABL33448;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1421.
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; neutropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX Homo sapiens.
XX WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP07537.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIC-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
XX Claim 1; SEQ ID NO 1421; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX Sequence 8842 BP; 2294 A; 186 C; 1955 G; 4405 T; 2 other;

Query Match 3.4%; Score 63.8; DB 24; Length 8842;
Best Local Similarity 44.9%; Pred. No. 0.037;
Matches 242; Conservative 0; Mismatches 297; Indels 0; Gaps 0;
QY 1340 ATTCTAGACAAAAATTTCAAAGGCTCTGAAAGTAAATGAAAAAGCGAGTTGTTTC 1399
DB 7578 AATCTTAATAAAAAAATTAATACTTAATCTACCAATATACCTTAATAATACAAACCA 7519
QY 1400 TATAGTAGAGCTGATCCCTGCTTAATGCTGTATATACACACTCTTTGGATAGATCC 1459
DB 7518 TAAACAAAAAATAATAGGTATATCTTCCCTCAATAAACCATTTATTAACAAAGC 7459
QY 1460 CAAAAAGAAACAAAAAATTTACCTTTGAAGATAGTGAATAAGAGAAACGTTTGTAGTCC 1519
DB 7458 AACAAACAAACAATATACATATCTAATAAATAAATAAATAAATAAATAAATAAATAA 7399
QY 1520 TCAGGTGTGTACAGACTAGAGAAAAAGTCTCAGTTTCCACCAATCCCAATTCAGATT 1579
DB 7398 TCAATTAACCTTACAAACAAACAATCCCACTAATATATCTCAAAAAATAAATAAATA 7339
QY 1580 TTAATTTTAAATTTGTAACACTGATATTTACTGCCAAATATAAGAAAAATTTTAAAGTA 1639
DB 7338 TAAAAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 7279
QY 1640 TGGTTATGTTGTAATTTTCAATGCTAATAGATAGTAGTCATATATATTTCAATT 1699
DB 7278 TAACAATCAACAATAATATATTTCTTACAAATTAATAAATAAATAAATAAATAAATA 7219
QY 1700 TCTTCATTACTTTAATGATTTTGTTCATGGCAGTTTGTAAAGTACTACTGATGATAT 1759
DB 7218 TTATTTATATATAAATAAATTTCCAAATACAACTACAAAAACTACAATTTTAAAAA 7159
QY 1760 TTTGCAATATTTATGTCACAGAAAAATTTTCATGTAAGTCATATTTTAAAGGAATAA 1819

Db 7158 AAAATCCAAACCAACCAAGACAATAACAAAAAATAAATCTTCTTCAATCCTA 7099
QY 1820 TACATAGCCCTTAAACAGTGTATTAACCTTTAAATGTAAAAAATAAATAAATAA 1878
Db 7098 AAATTAACAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7040

RESULT 11
AAS63334/C
ID AAS63334 standard; DNA; 8842 BP.
XX AC AAS63334;
XX DT
XX 29-JAN-2002 (first entry)
XX DE Chemically pretreated metabolism associated gene #29.
XX Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;
KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
KW single nucleotide polymorphism detection; SNP; stool; urine; lung;
KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;
KW DUSP2; EPHX2; QDPR; SGSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.
XX OS Homo sapiens.
XX PN WO200176451-A2.
XX 18-OCT-2001.
XX 06-APR-2001; 2001WO-EP04016.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-010834/01.
XX New nucleic acid, useful for diagnosis and therapy of metabolic
PT disease, solid tumour and cancers, comprises segment of chemically
PT modified genomic sequences of genes associated with metabolism -
XX
XX Claim 1: Page 89-91; 143pp; English.
XX The invention relates to a nucleic acid (I) comprising a sequence at
XX least 18 bases of a segment of the chemically pretreated DNA of genes
XX associated with metabolism such as DUSP2 (NM_004418), EPHX2 (NM_001979),
XX QDPR (NM_000320), SGSH (NM_000199), SHMT2 (NM_005412), SLC7A2
XX (NM_003046), SLC7A4 (NM_004173) and TYMS (NM_001071) (all
XX undefined). (I) are useful for diagnosis and therapy of metabolic
XX disease, solid tumours and cancers; as primer oligonucleotides for the
XX amplification of DNA sequences, for detecting the cytosine methylation
XX state and/or single nucleotide polymorphisms (SNPs) in a chemically
XX treated DNA of genes associated with metabolism. An array of (I) is
XX useful for ascertaining genetic and/or epigenetic parameters for the
XX diagnosis and/or therapy of existing diseases or the predisposition to
XX specific diseases by analysing cytosine methylations. The method involves
XX chemically treating genomic DNA sample by a solution of bisulphite,
XX hydrogen sulphite or disulphite such that cytosine bases which are
XX unmethylated at the 5th-position are converted to uracil or another base
XX which is dissimilar to cytosine in terms of hybridisation behaviour and
XX amplifying fragments of the chemically pretreated genomic DNA. The
XX genomic DNA is from cells or cellular components which contain DNA.
XX sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,
XX stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as
XX tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast
XX or liver, histologic object slides and their combinations. Genetic
XX parameters are mutations, in particular insertions, deletions, point
XX mutations, inversions and polymorphisms of genes associated with

CC metabolism and sequences further required for their regulation.
CC Epigenetic parameters are in particular cytosine methylations and
CC further chemical modifications of DNA bases of genes associated with
CC metabolism. Further epigenetic parameters include for e.g. the
CC acetylation of histones which correlates with DNA methylation.
CC AAS63306-AAS63373 represent chemically pretreated metabolism associated
CC genes, and related primers of the invention.
XX
SQ Sequence 8842 BP; 2294 A; 186 C; 1955 G; 4405 T; 2 other;
Query Match 3.4%; Score 63.8; DB 24; Length 8842;
Best Local Similarity 44.9%; Pred. No. 0.037;
Matches 242; Conservative 0; Mismatches 297; Indels 0; Gaps 0;
QY 1340 ATTCACTAGACAAAAATTTCAAAAAGGCTTGAAGTAAGTAAATGAAAAGCGAGTGTTC 1399
Db 7578 AATCTTAATAAAAAAATTAATACTATCTACCAAAATATACCCCTAAATAATACAAACAA 7519
QY 1400 TATAGTAGAAGCTGATCCCTGCTAATAATGCTGTATACACAACTCTTTGGATAGATCC 1459
Db 7518 TAAACAAAAAACAATTAACGTATATCTTCAACCCCTCAATAAACCAATTTTAAACAAACC 7459
QY 1460 CAAAGAAACAAAAAATTTACCTTTGAAGATAGTGAATAAGAGAAAAAGCTTTAGTGCC 1519
Db 7458 AACAAACACAAACAATAACAATACTAAAAAAAACAATAAATACTAAATAATCAATATT 7399
QY 1520 TCAGGTTGTCACAGACTAGAGAAAGTCTCAGTTTCACCAATCCCAATTCAAATGAGTT 1579
Db 7398 TCAATTAACCTTACAAACAAAAACAATCCACTAATATATCTCAAAAAAATAAAAAAAT 7339
QY 1580 TTAATTTTAATTTGTAATAAATCTGATTAATGCTCAATATATGAGAAATATTTTAACTAT 1639
Db 7338 TAAAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7279
QY 1640 TGGTTATGTTGTAATTTTCAATGTAATGCTAAATAGATAGTCAATATATATCAATT 1699
Db 7278 TAACATCAACAATAATATATTTCTTAACAATAAATAAATAAATAAATAAATAAATAA 7219
QY 1700 TCTTCATTAATTAATGTTGTTGTCATGGCAGTGTCTTAAAGTACATATCATGTGTATAT 1759
Db 7218 TTATTTATTTATAAATAAATAAATTTCCAAATACAACTACAAAAAATAAATAAATAA 7159
QY 1760 TTTGTCAATATTTATGTCACACAGAAATATTCATGTAAGTCATATTTTAAAGCAATAA 1819
Db 7158 AAAAAATCCAAACCAACCAAGACAATAAATAAATAAATAAATAAATAAATAAATAA 7099
QY 1820 TACATAGCCCTTAAACAGTGTATTAACCTTTAAATGTAAAAAATAAATAAATAA 1878
Db 7098 AAATTAACAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7040

RESULT 12
AAS46788/C
ID AAS46788 standard; DNA; 61020 BP.
XX AC AAS46788;
XX AC AAS46788;
XX DT 18-DEC-2001 (first entry)
XX DE Tumour suppressor gene derived chemically modified sequence #514.
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX OS Homo sapiens.
XX PN WO200168912-A2.
XX 20-SEP-2001.
XX 15-MAR-2001; 2001WO-EP02955.

PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2001-602752/68.
DR
XX Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer
XX
PS Claim 1; SEQ ID No 514; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (SS) and sequences complementary to (SS). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 other;

Query Match 3.4%; Score 63.4; DB 22; Length 61020;
Best Local Similarity 46.7%; Pred. No. 0.049;
Matches 276; Conservative 0; Mismatches 306; Indels 9; Gaps 2;

QY 1294 TAGTCGCTTATGGAACAACACTATATTCCACCTCTATTATATGCATTCAGTACAGCAA 1353
DB 49874 TAATAATCATCTTTAAAAAACAACAACTAACACCTATATCCCACTACTCCGTAAACCGA 49815
QY 1354 AATTTCAAAGCTGTGAAAGTAAATGAAAGCGAGTGTCTTATAGTAGAGCTG 1413
DB 49814 AATTAATAAATCACTTAAACCTTAAATAATTAACACTACAATAATCTATAATTACACTACT 49755
QY 1414 ATCCCTCGCTTAATGCTGTTAATACACAACCTCTTGATAGATCCCAAGAAACAAA 1473
DB 49754 ATACTACAACATAACAAACAAATAAACCCTATTTCAAAACAAACAAACAAACCAA 49695
QY 1474 AATATACCTTTGAAGAT---AGTGAATAAGAGAAAGAGTTAGTGCTCAGGTTGTCA 1530
DB 49694 ACAACACCGAATCATTTCTACTAAACAAACAACTTAAAAATCTAAAAATAAACHTACCTA 49635
QY 1531 CAGACTAGAGAAAGCTCAGTTTCACCAAAATCCCAATCAATGAGTTTAAATTTAAA 1590
DB 49634 CCAATATAACTAAACTTCAAAATTAATAAACAATAACATAAAAAAAATTAATAAATAAA 49575
QY 1591 TTGTAAAAACCTGATATTACTGCCAAATATAGAAAAATATTTAAGTATGTTATGTTG 1650

Db 49574 CAAAACAAATCACAATTAACCTATACAAAAATAAAAAATTTTAAACATCTTTTCCCTTAA 49515
QY 1651 TAAATTTTCAATGTGAAATGCTAATTAGATAGGTGCATATATATTCATATTCCTTCACTACT 1710
Db 49514 CATCTCTAATTAATAAACAACAATAATTTAAACACTAAAAAATTTAA-----AAATTA 49461
QY 1711 TAATGTATTTGTGCATGCGAGTGTGTTAAAGTACTATCATGTGTATATTTTGTGCAATAT 1770
Db 49460 CAAAAAATCATTAATCAAAATCTTTTAAATAAATTTTATATATATCCATTTTAAATCAATA 49401
QY 1771 TATGTCCAAACAGAAATATTCATGTAAGTCATATTTTAAAGCAATAAATACATAGCCTT 1830
Db 49400 TTTCTACAAATTTTCTCATTTTAAATTTCTACTCTCTAAATACCATACCTAAAAAT 49341
QY 1831 AAAACAGTGTATAACTTTTAAATGTAAAAATGAAAAAATAAAAAAATAAAAAA 1881
Db 49340 ACAACAATATAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 49290
RESULT 13
ABL32616/C
ID ABL32616 standard; DNA; 11555 BP.
XX
AC ABL32616;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 589.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-130909/17.
DR
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 589; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 11555 BP; 2890 A; 134 C; 2788 G; 5737 T; 6 other;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2002, 07:36:51 ; Search time 86 Seconds
(without alignments)
6707.663 Million cell updates/sec

Title: US-09-845-721-1
Perfect score: 1881
Sequence: 1 gttattttcctcaagga.....aaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1625	86.4	2238	2	US-08-919-624-2
2	60.2	3.2	5852	1	US-07-867-106-2
3	54.4	2.9	5852	1	US-07-867-106-2
4	54.4	2.9	19124	2	US-08-487-826B-13
5	53	2.8	1736	3	US-09-182-816-22
6	53	2.8	1736	3	US-09-182-816-24
7	53	2.8	1736	3	US-09-471-528-22
8	53	2.8	1736	3	US-09-471-528-24
9	53	2.8	1736	4	US-09-634-530-22
10	53	2.8	1736	4	US-09-634-530-24
11	51.6	2.7	20674	4	US-09-841-638-651
12	51	2.7	7218	1	US-08-232-463-14
13	50	2.7	1841	5	PCT-US95-00362-1
14	49.8	2.6	731	1	US-08-451-405A-2
15	49.2	2.6	1931	3	US-09-019-942-2
16	49.2	2.6	1931	4	US-09-099-041A-1
17	49.2	2.6	1931	4	US-09-245-281-1
18	49.2	2.6	1931	4	US-09-470-271-2
19	49.2	2.6	1931	4	US-09-207-359B-1
20	49	2.6	20674	4	US-09-641-638-651
21	48.8	2.6	828	4	US-08-998-416-538
22	48.2	2.6	1378	4	US-09-149-476-208
23	48.2	2.6	2323	4	US-09-149-476-24
24	48.2	2.6	2610	4	US-09-545-814-1
25	48.2	2.6	2610	4	US-09-545-814-3
26	48.2	2.6	3138	1	US-07-867-106-4
27	48.2	2.6	3527	2	US-08-909-965C-7

28 47.8 2.5 665 2 US-08-883-795A-36
29 47.8 2.5 3600 3 US-08-894-731-1
30 47.6 2.5 319 1 US-07-593-657-14
31 47.6 2.5 1241 1 US-07-593-657-6
32 47.6 2.5 1241 4 US-08-942-012B-3
33 47.6 2.5 2060 1 US-07-721-761A-31
34 47.6 2.5 2060 1 US-07-978-687-31
35 47.6 2.5 2060 5 PCT-US91-01750-2
36 47.6 2.5 2060 5 PCT-US91-05801-31
37 47.6 2.5 2081 5 PCT-US91-01750-3
38 47.6 2.5 3440 1 US-08-471-791-27
39 47.6 2.5 3440 5 PCT-US91-01746-27
40 47.4 2.5 4140 3 US-08-894-731-2
41 47.4 2.5 168575 4 US-09-426-290-1
42 47 2.5 1582 3 US-08-545-196B-10
43 47 2.5 1582 3 US-08-545-196B-12
44 46.8 2.5 1821 4 US-09-149-476-90
45 46.4 2.5 1117 4 US-09-247-373B-33

ALIGNMENTS

RESULT 1

US-08-919-624-2
; Sequence 2, Application US/08919624
; Patent No. 5994097
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: NEW HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,624
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0377 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: CARDNOT01
; CLONE: 282414
US-08-919-624-2

Query Match 86.4%; Score 1625; DB 2; Length 2238;
Best Local Similarity 97.1%; Pred. No. 0;

Matches 1677; Conservative 0; Mismatches 45; Indels 5; Gaps 2;

QY 139 TCTTAAAGAACACAGCTTATACGTTCATTTAAATTCGCAAAATATCAAAATAGTTTATTTCTAT 198
D 28 TATGGAGAGAAACGAAAGAAATCTCCAGACATGAAATAAACACAGACTACTTC --- 84
QY 199 TTCACCTTCTAGGGAAAAAACCAACTGCTCCAAAAGAAATGTTTTCTCCCATCTGG 258
D 85 -ACTGTGTCAGGGAAAAAACCAACTGCTCCAAAAGAAATGTTTTCTCCCATCTGG 143
QY 259 AAATCAACATGACGTGAATCTAATCAATTAACATTAACAGTGGAGATGACATTTGATGACATCAACA 318
D 144 AAATCAACATGACGTGAATCTAATCAATTAACATTAACAGTGGAGATGACATTTGATGACATCAACA 203
QY 319 CCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAGTGCTCTCACCGGATTC 378
D 204 CCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAGTGCTCTCACCGGATTC 263
QY 379 TTATGTTAGAAAATTTGTTGGAGCTTGGCAGCAACCTCACTGTATTTGGTACTTTACTGCA 438
D 264 TTATGTTAGAAAATTTTGTGACTTGGCAGCAACCTCACTGTATTTGGTACTTTACTGCA 323
QY 439 TGAATCCCAACTTAATCAACTCTGTCACTGTAACATTAATACATGAATCTTCACTGTACTTG 498
D 324 TGAATCCCAACTTAATCAACTCTGTCACTGTAACATTAATACATGAATCTTCACTGTACTTG 383
QY 499 ATGTAATATTTGTGGGATGATATCCCTCAACTATAGTTATCCCTTCTGCTTCACTGG 558
D 384 ATGTAATATTTGTGGGATGATATCCCTCAACTATAGTTATCCCTTCTGCTTCACTGG 443
QY 559 AGAGTAACACTGCTCATTTCTGTTTCCATGAGGCTTGTAATCTTTTGAAGTGCT 618
D 444 AGAGTAACACTGCTCATTTCTGTTTCCATGAGGCTTGTAATCTTTTGAAGTGCT 503
QY 619 CAACAGCAATCAACGTTTTTGTATCACTTTGGACAGATATCACATCTCTGTAACACCTG 678
D 504 CAACAGCAATCAACGTTTTTGTATCACTTTGGACAGATATCACATCTCTGTAACACCTG 563
QY 679 CAACCGAATTCGACATGGCAGAGCTGTAATGTTAATGATATCCATTTTGGATTTTT 738
D 564 CAACCGAATTCGACATGGCAGAGCTGTAATGTTAATGATATCCATTTTGGATTTTT 623
QY 739 CTTTTTCTCTTCTGATTCCTTTTATGAGGTAAATTTTTCAGTCTTCAAGTGGA 798
D 624 CTTTTTCTCTTCTGATTCCTTTTATGAGGTAAATTTTTCAGTCTTCAAGTGGA 683
QY 799 ATACCTGGGAAAAACAGACACATTTTATGTCAGTACAAATGAATACATCACTGAACCTG 858
D 684 ATACCTGGGAAAAACAGACACATTTTATGTCAGTACAAATGAATACATCACTGAACCTG 743
QY 859 GAATGTATTATCACCTGTTAGTACAGATCCCAATATTTCTTCTACTGTTGTAGTAATGT 918
D 744 GAATGTATTATCACCTGTTAGTACAGATCCCAATATTTCTTCTACTGTTGTAGTAATGT 803
QY 919 TAATCACATACACAAAATACCTTACAGCTCTTAATATTGGAATAGGCACAAAGATTCAA 978
D 804 TAATCACATACACAAAATACCTTACAGCTCTTAATATTGGAATAGGCACAAAGATTCAA 863
QY 979 CAGGCAGAGAAGAAAGCAAGAAAGAAAACAAATTTCTTACACACACACATGAGG 1038
D 864 CAGGCAGAGAAGAAAGCAAGAAAGAAAACAAATTTCTTACACACACACATGAGG 923
QY 1039 CTACAGACATGTACAAAACAGCTGGTGGGAGAAATGTAGTCTTTGGTGTGAAGACTTCAG 1098
D 924 CTACAGACATGTACAAAACAGCTGGTGGGAGAAATGTAGTCTTTGGTGTGAAGACTTCAG 983
QY 1099 TTTCTGTAATAATTCGCTCCGGCAGCTGTGAACAGCACCGTGAACGACGAGAAAGAC 1158
D 984 TTTCTGTAATAATTCGCTCCGGCAGCTGTGAACAGCACCGTGAACGACGAGAAAGAC 1043
QY 1159 AAAGAGAGCTCTTCAGGATGCTTTATTCATTTATTTCTACATTTCTCTCTCTGGACAC 1218
D 1044 TAAGAGAGCTCTTCAGGATGCTTTATTCATTTATTTCTACATTTCTCTCTCTGGACAC 1103

QY 1219 CAATTTCTGTTTTAAATACCACCAATTTATGTTTAGGCCCAAGTGACCTTTTAGTAAAT 1278
D 1104 CAATTTCTGTTTTAAATACCACCAATTTATGTTTAGGCCCAAGTGACCTTTTAGTAAAT 1163
QY 1279 TAAGATTTGTTTTTTTAGTCATGGCTTATGGAACAACATATATTACCCCTCTATTATATG 1338
D 1164 TAAGATTTGTTTTTTTAGTCATGGCTTATGGAACAACATATATTACCCCTCTATTATATG 1223
QY 1339 CATTCTACAGACAAAATTTTCAAAAGGCTCTGAAAAGTAAATGAAAAGCCGAGTTGTT 1398
D 1224 CATTCTACAGACAAAATTTTCAAAAGGCTCTGAAAAGTAAATGAAAAGCCGAGTTGTT 1283
QY 1399 CTATAGTAGAAGCTGATCCCTGCTCTAATAATGCTGTAATACACAACCTCTTGATAGATC 1458
D 1284 CTATAGTAGAAGCTGATCCCTGCTCTAATAATGCTGTAATACACAACCTCTTGATAGATC 1343
QY 1459 CCAAAGAAAACAAAATTTACCTTTTGAAGATAGTAGGAATAAGAGAAAACGTTTAGTGC 1518
D 1344 CTAAGAAAACAAAATTTACCTTTTGAAGATAGTAGGAATAAGAGAAAATCTTTAGTGC 1403
QY 1519 CTCAGGTTGTCACAGACTAGAGAAAAGTCTCAGTTTCACCAATCCACATTCANAATGAGT 1578
D 1404 CTCAGGTTGTCACAGACTAGAGAAAAGTCTCAGTTTCACCAATCCACATTCANAATGAGT 1463
QY 1579 TTTAAATTTAAATTTGTAATACTGATATTACTGCCAAATATAGAAAATATTTTAAAGTA 1638
D 1464 TTTAAATTTAAATTTGTAATACTGATATTACTGCCAAATATAGAAAATATTTTAAAGTA 1523
QY 1639 TTGGTTATGTTGTAATTTTCAATGTGAATGCTAATTAGATAGGTGATATATTCAAT 1698
D 1524 TTGGTTATGTTGTAATTTTCAATGTG-AAATGCAATTAGATAGGTGATATATTCAAT 1582
QY 1699 TTCTTCATTTACTTAATGATTTTGGCAGTGGCAGTTGTTAAAGTACTATCATGTATATA 1758
D 1583 TTCTTCATTTACTTAATGATTTTGGCAGTGGCAGTTGTTAAAGTACTATCATGTATATA 1642
QY 1759 TTTTGTCAATATTATGTCACACAGAAAATATTATCATGTAAGTCAATATTTTAAAGGAATAA 1818
D 1643 TTTTGTCAATATTATGTCACACAGAAAATATTATCATGTAAGTCAATATTTTAAAGGAATAA 1702
QY 1819 ATACATAGCTTTAAACACAGTGTATTAACCTTTAAATGTAATAAAAAAAA 1865
D 1703 ATACATAGCTTTAAACACAGTGTATTAACCTTTAAATGTAATAAAAAATGTAAGTGTATATA 1749

RESULT 2

US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526

GENERAL INFORMATION:

; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:

[illegible]

Db 5599 TAAA 5596

RESULT 4

US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487.826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NTH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

Query Match 2.9%; Score 54.4; DB 2; Length 19124;
Best Local Similarity 46.7%; Pred. NO. 0.019;
Matches 209; Conservative 0; Mismatches 236; Indels 3; Gaps 1;
QY 1437 ATACACAACTCTGTGATAGATCCCAAAAGACAAACAAATTTACCTTTGAAGATAGTGA 1496
Db 15472 ATAAAAAATTTTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15531
QY 1497 ATAAGAGAAACGTTTAGTGCTCAGGTGTCACAGACTAGAGAAAGTCTCAGTTTCA 1556
Db 15532 AAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15591
QY 1557 CCAATCCACATCAATGAGTTTAAATTTAAATTTGTAACATGATATTAATCTGCCAAA 1616
Db 15592 TCAGAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15651
QY 1617 TATAAGAAAAATTTTAAAGTATTTGGTATGTTGTAATTTTCAATGTGAATGCTAAT 1676
Db 15652 ACAAGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15711
QY 1677 AGATAGGTCATATATTCATTTTC---TTCATTACTTAATGATATTTGTTGATGGCAGT 1733

Db 15712 AAAGAAAAAATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15771
QY 1734 TTGTTAAAGTACTATCATGTGTATATTTTGTCATATTTATGTCACACAGAAATATTTCAT 1793
Db 15772 ATGTTAAAAAATAAATAATATATACATAAATAAATAAATAAATAAATAAATAAATAA 15831
QY 1794 GTAAGTCATATTTTAAAGGATAAATACATAGCCTTAAAAACAGTGTATACCTTTAAAAAT 1853
Db 15832 TAAATAAAAAATTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15891
QY 1854 GTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1881
Db 15892 AAAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 15919

RESULT 5

US-09-182-816-22
; Sequence 22, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182.816
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-182-816-22

Query Match 2.8%; Score 53; DB 3; Length 1736;
Best Local Similarity 59.7%; Pred. NO. 0.024;
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1733 TTTGTTAAAGTACTATCATGTGTATATTTTGTCATATATATGTCACACAGAAATATTTC 1792
Db 1574 TTTGTGATAATAATAATATGTTAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1633
QY 1793 TGTAAAGTCATATTTTAAAGGATAAATAACATACATAGCCTTAAAAACAGTGTATACCTTTAAAA 1852
Db 1634 TTTTATTTCAACTGTCATATATATAAATAAATAAATAAATAAATAAATAAATAAATAA 1693
QY 1853 TGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1881
Db 1694 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1722

RESULT 6

US-09-182-816-24/c
; Sequence 24, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182.816
; CURRENT FILING DATE: 1998-10-29

; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-182-816-24

Query Match 2.8%; Score 53; DB 3; Length 1736;
Best Local Similarity 59.7%; Pred. No. 0.024;
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 1733 TTTGTTAAAGTACTATCATGCTGATATTTTGTCAATATTTGTCACACAGAAATATTC 1792
DB 163 TTTGTGATAATAATATATGTTAAATAATATGTAATCTGTGAATAAACGATATGGA 104
QY 1793 TGTAAAGTCATATTTTAAAGCAATAATACATAGCCCTTAAACAGTGTATTAACCTTTAAAA 1852
DB 103 TTTTATTCAAACTGTCAATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 44
QY 1853 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1881
DB 43 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 15

RESULT 7

US-09-471-528-22
; Sequence 22, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-471-528-22

Query Match 2.8%; Score 53; DB 3; Length 1736;
Best Local Similarity 59.7%; Pred. No. 0.024;
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 1733 TTTGTTAAAGTACTATCATGCTGATATTTTGTCAATATTTGTCACACAGAAATATTC 1792
DB 1574 TTTGTGATAATAATATGTTAAATAATATGTAATCTGTGAATAAACGATATGGA 1633
QY 1793 TGTAAAGTCATATTTTAAAGCAATAATACATAGCCCTTAAACAGTGTATTAACCTTTAAAA 1852
DB 1634 TTTTATTCAAACTGTCAATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1693
QY 1853 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1881
DB 1694 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1722

RESULT 8

US-09-471-528-24/C
; Sequence 24, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-471-528-24

Query Match 2.8%; Score 53; DB 3; Length 1736;
Best Local Similarity 59.7%; Pred. No. 0.024;
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 1733 TTTGTTAAAGTACTATCATGCTGATATTTTGTCAATATTTGTCACACAGAAATATTC 1792
DB 163 TTTGTGATAATAATATGTTAAATAATATGTAATCTGTGAATAAACGATATGGA 104
QY 1793 TGTAAAGTCATATTTTAAAGCAATAATACATAGCCCTTAAACAGTGTATTAACCTTTAAAA 1852
DB 103 TTTTATTCAAACTGTCAATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 44
QY 1853 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1881
DB 43 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 15

RESULT 9

US-09-634-530-22
; Sequence 22, Application US/09634530
; Patent No. 6290958
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/634,530
; CURRENT FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 09/471,528
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 09/182,816
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 08/989,510
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-634-530-22

Query Match 2.8%; Score 53; DB 4; Length 1736;
Best Local Similarity 59.7%; Pred. No. 0.024;

NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTgpt-F1s
US-08-232-463-14

Query Match 2.7%; Score 51; DB 1; Length 7218;
Best Local Similarity 12.7%; Pred. No. 0.078;
Matches 48; Conservative 167; Mismatches 162; Indels 0; Gaps 0;
QY 333 CCACATATCATCCGTTAAGCTTTCAAGTGTCTCTCACCGGATTTCTTATGTAGAAAT 392
Db 1149 YY 1208
QY 393 GTGTTGGGACTTGGCAGCAACCTCACTGTATTGGTACTTACTGTCATGAAATCCAACTTA 452
Db 1209 YY 1268
QY 453 ATCAACTCTGTCTCAGTAACATATTACAATGAATCTTCAATGACTTGTATGTAATTTGT 512
Db 1269 YY 1328
QY 513 GTGGGATGATTCCTTAAGTATAGTATCTCTGCTTCTACTGGAGATCAACACTGCT 572
Db 1329 YY 1388
QY 573 CTCATTGCTGTTCCATCAGCTTGTGTATCTTTTGCAGTGTCTCACACCAATCAAC 632
Db 1389 YY 1448
QY 633 GTTTTGTCTATCATTGTCAGAGATATGACATCTCTGTAACACCTGCAACCGAATTCG 692
Db 1449 TATCTCTTAACACTTGTGATAGTAGTAAATACAGTATGCTTACATGCGCTTTTGT 1508
QY 693 ACAATGGGCGAGCTGT 709
Db 1509 AAACGAATAGATGCT 1525

RESULT 13
PCT-US95-00362-1
; Sequence 1, Application PC/TUS9500362
; GENERAL INFORMATION:
; APPLICANT: IMMUNEX CORPORATION
; TITLE OF INVENTION: Ligand That Binds Fas Antigen
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00362
; FILING DATE: 06-JAN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,138
; FILING DATE: 07-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/190,559
; FILING DATE: 01-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Anderson, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2805-WO
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 93..938
; PCT-US95-00362-1
Query Match 2.7%; Score 50; DB 5; Length 1841;
Best Local Similarity 53.9%; Pred. No. 0.099;
Matches 125; Conservative 0; Mismatches 105; Indels 2; Gaps 1;
QY 1652 AAATTTCAATGTGAATGCTAATTAGATAGGTGCATATATATATATTTCTTCAATCTT 1711
Db 1592 ATATTGTGAAGTACATATTAGGAAATATGGCTTGCATTTGCTCAAGATTTTGAATGCTT 1651
QY 1712 AATGATTGTTGTCATGGCGATTTGTTAAAGTACT--ATCATGTGTATATTTTGTCAATA 1769
Db 1652 CCGTCAATCAACTCTAATAGTGTGTTAAATCAATTCATTGTTGCTACTAATGATGTTT 1711
QY 1770 TTATGTCCCAACAGAAATATTCATGTAGTCATATTTTAAAGGAATAATACATAGCCT 1829
Db 1712 TCTTATATATTAATTAATATTTATGTAGTGTGCAATTTTGTGAATGAAACATGTAAT 1771
QY 1830 TAAACAGTGTATACTTTTAAATGTAAAAAATGTAATGTAATGTAATGTAATGTAATGTA 1881
Db 1772 AAAAAGTATATCTAGGATACAAATAAAAAAATGTAATGTAATGTAATGTAATGTAATGTA 1823
RESULT 14
US-08-451-405A-2
; Sequence 2, Application US/08451405A
; Patent No. 5736358
; GENERAL INFORMATION:
; APPLICANT: FASEL, NICOLAS JOSEPH
; REYMOND, CHRISTOPHE DOMINIQUE

;; TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
;; TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
;; NUMBER OF SEQUENCES: 3
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: THE WEBB LAW FIRM
;; STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
;; CITY: PITTSBURGH
;; STATE: PENNSYLVANIA
;; COUNTRY: UNITED STATES OF AMERICA
;; ZIP: 15219-1818
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5" FLOPPY DISK
;; COMPUTER: Midwest Micro 486-50
;; OPERATING SYSTEM: DOS
;; SOFTWARE: WORDPERFECT 6.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/451,405A
;; FILING DATE: 26-MAY-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/965,273
;; FILING DATE: 15-JAN-1993
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 731
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: SINGLE
;; TOPOLOGY: UNKNOWN
US-08-451-405A-2

Query Match 2.6%; Score 49.8; DB 1; Length 731;
Best Local Similarity 45.4%; Pred. No. 0.092;
Matches 303; Conservative 0; Mismatches 352; Indels 12; Gaps 3;
QY 1220 AATTCTGTTTTAAATACCACCACTTTATGTTTAGGCCCAAGTGACCTTTTAGTAAAT 1279
DB 58 AGTATATGTTTTTTCCTTTTTTTTATTTATTTTATTTTAAATAATAAAT 117
QY 1280 AGATTGTTTTTTTAGTCATGCGCTTATGGAACAATATATTCACCTCTATATATGC 1339
DB 118 TAGAATAAATAATATCTTATTTGAAGGAGTTTTTATTTGTTATTTAAATATATATTAACAT 177
QY 1340 ATTCA--CTAGACAAAATTTCAAAGGCTTCGAAAGTAAAA--TGAAAAAGCGAGTT 1394
DB 178 AGTGAACCTAATAATAGATTGTGACGGTATATGTAAGAAATTCATAAAAAAATTC 237
QY 1395 GTTCTATAGTAGAAGCTGATCCCTGCTTAATGCTGTATACACAACTCTTGATA 1454
DB 238 AGATAATTTTGCATTGGAACAACAACAACAAAAAATAAAAAAATAAAAAATC 297
QY 1455 GATCCCAAGAACAACAAAAAATTTACCTTTGAAGATAGTGAATAAGAGAAAAACGTTA 1514
DB 298 AAAAAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 357
QY 1515 GTCCCTCAGTTGTCACAGACTAGAGAAAGTCTCAGTTTCACCAATCCACATTCAAAT 1574
DB 358 ATATTATATATATATATTTTAAATGCGAAACACACTTTTAAACACACTCTATTATCTACA 417
QY 1575 GAGTTTTTAAATTTGTAATAAAGCTGATTTACTGCCAAATATAGAAAAATATTTTA 1634
DB 418 AAGGTTTTAAATTTTAAATTTTATTTTATTTATTTTATTTTAAATAAATTTT 477
QY 1635 AGTATGTTGTTATGTTGTAATTTTCAATGTAATGTAATGTAATGTAATGTAATGTAAT 1694
DB 478 ATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 530
QY 1695 CAATTCTCTCATTTAATGTTATTTGTCATGGCAGTTTGTGTAAGTACTATCATGTG 1754
DB 531 ACAATTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 590
QY 1755 TATATTTTGTCAATATTTATGTCACAGAAAAATTTATCATGTAAGTCTATATTTTAAAGGA 1814
DB 591 TAGTAATAATAATTTCTTTTTTAAATAAATAACCTTTTTTCTACATTTATTTTATTTATTTA 650

QY 1815 ATAAATACATAGCCTTAAACACAGTGTATAACTTTAAATCTAAAAATCTAAAAA 1874
DB 651 TTTTCTTTCTTAATCATTCATAAATTTTATTTTAAAAAATAAATAAATAAATAA 710
QY 1875 AAAAAA 1881
DB 711 ACAACA 717
RESULT 15
US-09-019-942-2
; Sequence 2, Application US/09019942
; Patent No. 6033855
; GENERAL INFORMATION:
; APPLICANT: Berlin, John
; TITLE OF INVENTION: GENES ENCODING CASPASE RECRUITMENT
; TITLE OF INVENTION: DOMAIN POLYPEPTIDES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,942
; FILING DATE: 06-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/068001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-019-942-2

Query Match 2.6%; Score 49.2; DB 3; Length 1931;
Best Local Similarity 65.5%; Pred. No. 0.14; Mismatches 0; Gaps 0;
Matches 72; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1772 ATGTCACACAGAAAATATTCATGTAGTCAATATTTTAAAGGAATAATACATAGCCTTA 1831
DB 1811 ATTTCTTCAATAAATAAAGCATGTAGTCAAGGAATAATGTGTTTCATAA 1870
QY 1832 AACAGTGTATACCTTTAAATGTAAAAAATAAATAAATAAATAAATAAATAA 1881
DB 1871 AAGGATATTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1920

Search completed: December 10, 2002, 09:46:51
Job time : 161 secs

This Page Blank (uspto)